

The long shadow from Chernobyl

Last week's reactor accident in the Soviet Union is an historic event that will change the course of the remainder of the twentieth century. But the clouds may also have a silver lining.

EUROPEAN governments' attempts to deal with the consequence of the reactor accident at Chernobyl have been seriously impeded by the lack of data from the Soviet Union. Information bearing on two crucial issues has been withheld. At least for the first week after the accident, the Soviet side had nothing to say about the composition of the radioactive material from the damaged reactor. And there has been no serious attempt to estimate what quantities of radioactivity were released. Even allowing for the difficulty of making considered judgement while the graphite moderator of the reactor was still on fire, for the need to protect the local population from serious long-term consequences and for national pride, there is no excuse for having kept silent on these issues for so long.

That is why the Soviet Union may have only itself to blame for the indecently gleeful way in which Western observers of this tragic accident, especially some in the United States, have given currency to wild speculations about the scale of the damage near Chernobyl. In reality, there is no reason to disbelieve the Soviet statement that only two people were killed by the explosion that appears to have marked the onset of the release of radioactivity; the more important questions are whether local fallout has caused potentially fatal radiation sickness among the neighbouring population, and what will be the longer-term load of increased cancer in a wider population.

When the dust has almost literally settled, there will be other international issues on which a fuller degree of Soviet understanding of other people's problems will be required. Who will meet the cost of the trouble which the Scandinavian governments have had to take? What if events in the next few months suggest that some people outside the Soviet Union have been acutely damaged? And what happens twenty or thirty years from now if people die from cancers which may be caused by radioactive effluent from the Chernobyl reactor, but which are most probably due to natural causes? Will those affected have a suit against the government of the Soviet Union? Or merely grounds for a lasting grudge? On past form, the Soviet Union will be grudging with helpful information, perhaps because it is unwilling to admit to its own people that its domestic arrangements are not by definition unalloyed benefits. On this occasion, the pressure for openness from outside may happily coincide with what Mr Mikhail Gorbachev has been advocating domestically.

The consequences of the accident for Soviet plans for the further development of nuclear power will nevertheless be substantial. It is unthinkable that the Soviet government will push forward with the construction of reactors of the type that caught fire at Chernobyl until it fully understands the risks of a repetition of the accident. The essential weakness of the design is its assumption that two mutually energetically reactive materials (graphite and steam) can be rigorously separated from each other by means of welded steel tubes; the assumptions are obviously no better than the quality of the welds. The obvious difficulty now for the Soviet government is that its plans for the generation of electricity hang crucially on the construction of nuclear reactors. Luckily, most of the plants now being built are the more rigorously tested pressurized-water reactors. But it will

serve nobody's interests if the Soviet economy, and its refreshing new leader, are set back a decade or so in their expectations for the years ahead.

The consequences of Chernobyl for the nuclear industry elsewhere will be devastating. Mr Saul Steinberg, the US financier, has made something of a splash in the past few weeks by buying a substantial proportion of the shares of the Long Island Lighting Company, the owner of the Shoreham nuclear power plant which has lacked an operating licence for the past several years. His envious competitors had been admiring his flair for making wagers that cannot fail; either Shoreham would be allowed to produce electricity, or the utility that owns it would be nationalized (by the state of New York). But now, Mr Steinberg will be seen to be betting on a single horse, one that hobbles a little. In the adversarial process by which the public good is balanced against the private interest in the United States, the dice will now be further weighted against those who may wish to build the odd nuclear reactor. Even in Britain, where the government is likely to be told in September by the Roskill Commission that it would be safe to allow the nationalized electricity utility to build a pair of pressurized water reactors, there will be a temptation to postpone the decision. In short, whatever the explanation of the Chernobyl accident, the event will reinforce the general prejudice against even the sensible use of civil nuclear power.

The trouble is the public understanding of the concept of probability. Nobody disputes that a substantial release of radioactivity into the atmosphere can damage people and that reactors are a potential source of large quantities of activity. Most designers of nuclear power stations have taken to calculating the chance that the plants they build will suffer a catastrophic accident; the usual outcome is that the chance of catastrophe is one in a million or more per year of operation. Put simply, reactors would not be built if the chance that they would go seriously wrong were much greater, for then utility systems operating several reactors would imply to their taxpayers that there is a chance that a catastrophe would arise within a human lifespan or thereabouts. But this average time does not exclude much shorter intervals, or excuse carelessness.

What side-benefits can there be in such a sombre litany of trouble? Much depends on how serious the local consequences of the accident have been. But even if, as must be hoped, there is no serious loss of life in the immediate neighbourhood of the reactor, there is bound to be a moral obligation on the Soviet Union to acknowledge the trouble and even damage it has caused elsewhere. For what it is worth, the Soviet Union is not a signatory of the 1963 Vienna Convention on the regulation of nuclear accidents, but the past few years have shown clearly that international conventions have a moral force that non-signatories cannot shrug off. Similarly, within the Soviet Union, there is bound now to be built-in pressure on the authorities to be more open about the risks of nuclear power, and of some of the other schemes on which the Soviet Union is engaged. And of course, if the accident has been as serious as it might have been, the internal pressures may be even greater — perhaps even greater than those needed to blow the lid off an isolated nuclear reactor. □



Chernobyl

Anxiety about reactor accident subsides

LAST week's panic in Western Europe and elsewhere about the explosion and subsequent graphite fire at the Soviet nuclear plant north of Kiev was in retreat earlier this week, as the scanty data available outside the Soviet Union were pieced together. While it is plain that the accident at Chernobyl on 26 April was the world's most serious nuclear accident so far, there is no reason to dispute the Soviet statement that only two people at the plant were killed, perhaps a consequence of the initial explosion. The longer-term consequences, including cases of radiation sickness which will become apparent in the next few weeks, and excess cancer deaths, which will be delayed by decades, cannot yet be estimated.

Part of the basis for this reassuring conclusion is the examination of members of the party of British students from Kiev and Minsk, who returned via Moscow last week. Although some of the party's clothing was contaminated with local fallout, ingestion of airborne radioactivity was apparently not sufficient to cause risk of thyroid cancer. An account of the fallout over Poland during the past week, given by a group of Polish experts in Vienna, has led the Polish authorities to the conclusion that there is "no significant risk to the health of any person".

Other estimates of the total quantity of radioactivity released from the Chernobyl reactor have been derived from computer models normally used in the assessment of the risks of nuclear accidents. These suggest that the release from the plant may have been upwards of two orders of magnitude greater than from the British accident at Windscale in 1957, previously the worst nuclear accident on record.

The Swedish measurements, the first clue (on 28 April) to the accident at Chernobyl, suggest that airborne activity over Scandinavia peaked on Tuesday last week at about 10 to 20 times the normal level, mostly because of iodine-131 and caesium-137.

Fallout in the neighbourhood of the plant has been much more serious. Close on 50,000 people have been evacuated from a 30 km radius. Soviet spokesmen visiting the West have spoken of radiation doses at the plant which were initially 200 roentgens an hour, falling to 100 roentgens an hour after a week. In this light, the Soviet statement that 200 people were hospitalized (of whom a quarter were discharged) probably refers to those showing symptoms of radiation sickness.

The cause of the accident remains a

matter for speculation. The most common suggestion is that a leak in the steam-raising water circuit which contains the fuel elements and interpenetrates the 1,000-tonne graphite moderator would have produced water gas (carbon monoxide and hydrogen), which might then have caused the explosion that destroyed the roof of the reactor building. It is not known whether Soviet statements that "human error" was responsible for the accident refers to the execution of an improper operation or to an inadequate response to warning signals.

Another possibility is that the accident arose during an operation to anneal the graphite moderator in the reactor to remove structural defects accumulated during prolonged neutron irradiation, and which entails heating the graphite above the normal operating temperature for more than a day.

Although the Chernobyl accident may have been less serious than first reports in Western Europe suggested, most Western governments have been indignant that so little information has been available from Soviet sources to suggest how governments elsewhere might best deal with the fallout from the plant. More detailed information about conditions at Chernobyl began appearing only earlier this week. On Tuesday *Pravda* carried a long article describing conditions around the reactor site.

Perhaps more significantly, an apparently painful interview last Wednesday between the director-general of the International Atomic Energy Agency, Hans Blix, and the Soviet envoy in Vienna had led, by Sunday, to an invitation for Blix to visit Moscow. He left the same day, accompanied by Maurice Rosen, the agency director of nuclear safety.

According to a spokesman of the agency on Tuesday this week, the intention is that the visitors will have full access to what information there is available in Moscow, although it is not known whether the party will be able to visit the reactor site. Blix's companions may stay in the Soviet Union for several days.

Within the Soviet Union, a commission of enquiry under a deputy prime minister is said to be hard at work on an investigation of the causes of the accident and into its aftermath. On the assumption that the reactor building can itself eventually be made safe, the most haunting problem will be that of managing the health of the population in the neighbourhood of the accident over the next three decades. □

Soviet nuclear reactors

Ambitious plans with red tape

UNTIL 26 April, the Soviet Union had a nuclear generating capacity of 28 GW, 11 per cent of total electricity output, but the new five-year plan (1986-90) includes the commissioning of a further 41 GW of nuclear capacity.

Guarded comments by Soviet officials last week suggest that there will be no change in these plans, even though it has for the first time been admitted that the use of nuclear power, even for peaceful purposes, cannot be totally safe. Previously, the Soviet line has been that incidents such as Three Mile Island were the result of corner-cutting by capitalists bent on profit.

Soviet plans specifically suppose the construction of nuclear stations close to major cities. Although, a few years ago, a leading party monthly carried an article suggesting that nuclear stations should be confined to the remoter areas of Siberia, they are now considered so safe that they can serve as thermal generators for city district heating systems, but with a green belt of 1-3 km radius around each station, which would be used for recreational purposes and allotments.

Current plans call for several such "atomic-thermal-power" stations to be commissioned during the next five years, including Gor'kii, Voronezh, Odessa and Minsk, while the construction of several more is to be started. If the Chernobyl accident has any effect on Soviet nuclear policy, it is in this sphere that it is most likely to be felt.

The resting of at most 10 or 12 power stations (and the quiet dropping of the district heating part of the programme) would be far less difficult, both logistically and politically, than a total reversal of Soviet commitment to nuclear energy. There is likely, too, to be an increased impetus to the fusion energy research programme, which for years has been formally hailed as the ultimate solution to Soviet energy problems.

The Chernobyl station had four RBMK reactors with a fifth under construction. This is a light-water-cooled uranium-graphite channel system, with zirconium alloy fuel cans. As a number of commentators have noted, this type of design has been rejected in Western countries as unsafe, and the presence of the earliest example of this type near Leningrad has, during the past few days, evoked considerable concern in Finland.

The Soviet designers seem to have relied on the regular ultrasonic and acoustic monitoring of the metal coolant loops, which, it was claimed in the technical journal *Atomnaya Energiya* (50,

253; 1981) "guaranteed that sudden rupture of the pipes could not occur" and, as a measure of last resort, the incorporation of an automatic emergency water cooling system. This was designed to cope with the worst possible case foreseen, the transverse fracture of a 900-mm turbine pressure collector. On 26 April, however, this system clearly failed to work.

If the Soviet atomic energy commission preserves the reticence it has shown so far, it may be difficult to discover not only what caused the initial fault but also why the emergency system failed to operate. (Earlier this year, the daily *Sovetskaya Rossiya* called for greater openness in the reporting of natural disasters, but the atomic energy commission could well reply that the Chernobyl accident was not "natural".) One possible explanation, however, may lie in Soviet construction procedures which are prone to produce shoddy and hazardous results, in spite of the fact that there is no "capitalist" profit motive.

The Soviet practice of laying down quarterly, annual and five-year production targets (with pay and bonuses dependent on their fulfilment), coupled with a highly bureaucratic supply system and a transport network too overloaded to ensure prompt deliveries, has led to the practice of "storming" — last-minute speed-ups and overtime to catch the target date.

Hasty construction is further exacerbated by the emphasis placed on fulfilling targets ahead of plan, and, in particular, the making of "voluntary" pledges by the workers to reduce their target time in honour of major patriotic events. Although, in 1976, a glowing description of the construction work at the Chernobyl site merited front-page treatment in *Pravda*, an article that appeared in the weekly *Literaturna Ukrayina* just a month before the accident gave a very different picture.

According to the author, Lyubov Kovalevs'ka from the town of Prypyat', building standards have from the very beginning been neglected. The construction of a project such as a nuclear power station, she says, must be "a continuous flow on the basis of the strictest adherence to building technology. But this did not happen here."

The various problems that arose with the first unit (Kovalevs'ka does not specify what they were) were never properly solved, but simply "transferred" to the second, and then to the third and fourth units.

What this meant in practice is seen from her description of current construction work on the fifth unit. The project, she says, is beset by constant shortages; materials are delivered in inadequate quantities and are frequently sub-standard, yet the planned construction

time has been cut from three years to two. The workforce therefore has no option but to try to fulfil its targets by neglecting the specifications and "stretching" such materials as there are as far as possible, and accepting defective supplies even when so vital an item is concerned as the

Nuclear prognostications

Estimating gravity of problem

ALTHOUGH several agencies in the West have been hard at work during the past week trying to estimate the seriousness of the accident at the Soviet nuclear reactor north of Kiev, nobody is confident of the outcome of their calculations. As part of the now-standard process of calculating the consequences of hypothetical nuclear accidents, several computer models have been developed for estimating what happens to radioactivity released from a damaged reactor. But, as one modeller explained last week, "usually we start with something that we know".

Part of the problem is meteorological. One of the team at the British Radiological Protection Board, an independent government agency based at Harwell, Oxfordshire, explained last week that there is relatively little experience of even models of the spread of radioactive materials from sources remaining active for several days, and that the calculation of the consequences of Chernobyl have been complicated by the rapidly changing wind patterns.

Data from Sweden and elsewhere in Europe have thrown some light on the composition of the fallout from the Soviet reactor. It seems to be agreed that the concentrations of volatile materials such as radioactive iodine and caesium are in the proportions to be expected from fresh fission products. The modellers are, however, unable to estimate with any certainty what proportion of the total emission from the damaged plant will have been deposited in its immediate locality.

This uncertainty bears most directly on the difficulty of estimating the damage that may have been done to the population in the immediate vicinity of the plant, which will become apparent there only in the next few days and even weeks. One of those concerned says the uncertainties are such that the data from outside the Soviet Union are consistent both with the assumption that many people will have been exposed to damaging doses of radiation and that there will be no casualties from radiation sickness even in the neighbourhood of the plant.

There is, however, general scepticism about some of the estimates of Soviet casualties such as that given to a congressional committee by Dr Kenneth Adelman, director of the US Arms Control and Dis-

case she cites, the roof-plate for the spent fuel bunker. Quality and inspection of the finished construction moreover, seems to connive at the situation which, Kovalevs'ka warned a month ago, will have to be "paid for over the course of decades". Has this bill yet been paid? **Vera Rich**

armament Agency. It is pointed out that calculations of the consequences of releases of radioactivity from pressurized water reactors, such as the estimates prepared for the public inquiry into the British plan to build such a plant at Sizewell in Suffolk, suggest that acute deaths (within a few weeks) would be counted in hundreds, not thousands. The longer-term consequences — excess cancer deaths — would usually be greater in number.

The generalized effects of the radioactivity observed in Sweden will be much smaller. If atmospheric radioactivity was last week six times greater than the natural background for two or three days, the result will be increased exposure to radiation by a few per cent of the annual dose arising naturally. But this reassuring yardstick does not exclude the possibility that some people may ingest exceptional amounts of artificial radioactivity by the accumulation and concentration of particular radionuclides in foodstuffs.

Attempts to estimate the total amount of radioactivity released at Chernobyl are similarly uncertain. The source of the estimate that a total of 10^{16} bequerel were released at Chernobyl was most of all anxious that the uncertainties, which must be several orders of magnitude, should be clearly appreciated. By this yardstick, the Chernobyl accident may be 100 times potentially more damaging than that at Windscale in 1957, when an air-cooled graphite reactor caught fire. That accident was in turn much more serious, as measured by the total quantity of radioactivity released, than the accident at Three Mile Island eight years ago.

Among those with military and arms control interests, some importance is attached to the likelihood that reactors of the "Leningrad" type such as those at Chernobyl would have been a useful source of military plutonium.

According to the International Atomic Energy Agency's survey of the world's reactors published at the end of last year, there were 26 light-water-cooled graphite-moderated reactors in operation in the Soviet Union at the end of 1984, producing between them a total of 13.7 GW of electricity. A further nine reactors of this type were being built at that stage, although the bulk of the reactors now under construction in the Soviet Union are pressurized-water reactors. □

France in Tokyo

Two heads better than one?

FRANCE has had two heads at the economic summit in Tokyo, one of them gung-ho for technological cooperation, and the other apparently lukewarm. The future of French policy in this increasingly important area of politics, which covers matters from the Strategic Defense Initiative (SDI) to joint industrial development and the mobility of scientists, is thus becoming somewhat murky.

At previous summits, President François Mitterrand, head of state and a socialist, has blown the trumpet for technology. His ministers, until the March election which threw them out, energetically pursued at least European co-operation in high technology and research. He launched Eureka, a kind of European civil answer to SDI. From the new Prime Minister, the right-wing Jacques Chirac, who is accompanying Mitterrand in Tokyo, little has been heard on these matters, apart from his proposing a candidate for foreign minister who was highly favourable to French government participation in SDI. Mitterrand vetoed the appointment: he believes out-and-out French participation will lead to a brain-drain, and a loss of French control of important new industrial technologies.

If this was a negative step (in the sense of international cooperation in technology), Chirac has also taken negative steps in his major reduction of the French research budget (see *Nature* 321, 8; 1986) — which will reduce the capacity of French bodies to join international collaborations — and in a sweeping reduction in the powers of the research ministry, which has lost control of technology policy as well as a number of technological programmes (such as space and energy) to the traditionally inward-looking ministry of industry.

Meanwhile other major divisions seem likely to plague the house of French international cooperation. A highly competent official, Robert Chabbal, ex-director of

science for the North Atlantic Treaty Organisation (NATO), has been appointed coordinator for international research affairs, but he will reside in the weakened ministry of research. The ministry's links with the Elysée, President Mitterrand's palace and office across the Seine, were originally strong and well-used in the development of Eureka, but they must now seem somewhat redundant as both parties have lost influence.

The demise of the ministry of research was clear in the new government's attributions of power and budget: the decline of the Elysée is less easy to identify.

When before the election Mitterrand was discussing the possible "cohabitation", as the French like to call it, between himself and a potential right-wing prime minister, he made it clear that he would hold on to one thing: foreign policy. This area of action for the President would of course include his dreams of a "technological Europe", driven by France.

Since the election, however, Chirac, apart from losing his first choice of foreign minister, who might anyway have been a symbolic sacrifice, has made it clear that he, not the President, will determine French foreign policy. Chirac had made four foreign tours before the visit to Tokyo to establish his links with French Africa, and the British, German and Italian Prime Ministers, while President Mitterrand has remained at home. Moreover, Chirac has established a private foreign policy advisory unit in his offices in the Hôtel Matignon, and it seems clear now that French foreign policy is being run by Chirac and the foreign ministry at the Quai d'Orsay. Links between the Elysée, where the Eureka programme originated, and the ministry of research are therefore unlikely to have the importance they once did; in Tokyo, Mitterrand is little more than a constitutional monarch, and his past presidential pro-

Good week for Pasteur institute

THE Institut Pasteur, the biological and medical research institute in Paris founded by Louis Pasteur in the nineteenth century, has learned from a US court that it may, after all, be granted priority in its patent dispute with the US National Institutes of Health over who first invented blood tests for AIDS (acquired immune deficiency syndrome).

Millions of dollars could be involved in the argument, which dates back to September 1983, when the Pasteur filed a patent based on the discovery by its researcher Luc Montagnier of the virus that causes AIDS ("HIV", according to the recommended new nomenclature). Robert Gallo of NIH filed a similar patent based on his own work many months later, and was awarded his while the French application is still pending.

Now, the US Patents and Trademark Office has announced "an interference" between the applications, in terms which, according to the Pasteur, mean that the institute will be awarded the patent rights unless Dr Gallo can prove prior discovery of HIV. Here Pasteur scientists feel on firm ground, but recent statements by Gallo (see p. 119) may give cause for doubt.

Moreover, all is not yet jubilation in Paris, as a final judgement in the US courts will take two years, and any redistribution of royalties on tests will take place only from the date of the judgement and will not be retroactive. Meanwhile, attempts to achieve a negotiated end to the dispute continue, and at the Pasteur (already heartened by a bequest from the Duchess of Windsor) they now feel they have a trump card. Robert Walgate

nouncements on a new world of technology merely echoes.

What everyone is waiting for, however, is to see where the new power, Chirac, moves in this area. His new foreign policy advisers have a strong industrial slant. Their chief is François Bujon de l'Estang (not to be confused with ex-President Giscard d'Estaing), a past international relations director of the French atomic energy agency (CEA), head of the US subsidiary of the nuclear fuel company COGEMA and chief cabinet adviser to Giscard d'Estaing's minister of industry, André Giraud. Moreover, Chirac's ministry of industry is as strong as the ministry of research is now weak. However, with the appointment as his science adviser of Yves Durant, an advocate of the freedom of research from state interference, who will be counselling Chirac to reduce the power of the ministry of research and its organizations still further, the chance of a renewed campaign by France for a research-led, high-technology future seems small. Robert Walgate

ESF joins Ocean Drilling Program

Washington

THE European Science Foundation (ESF) has signed on as the seventh member of the Ocean Drilling Program (ODP), an international project for ocean basin exploration. ESF has been interested in joining for some time, but the \$2.5 million membership price tag was a stumbling block that has only recently been overcome.

As an ODP member, ESF can send two scientific representatives on each ODP cruise, and one co-chief scientist each year. ESF will also participate in mission

planning and site selection.

Representatives of ESF and the National Science Foundation signed the memorandum of understanding sealing the deal at a meeting in Annapolis, Maryland, last week of the ODP executive committee. ESF joins France, West Germany, the United Kingdom, Japan, Canada and the United States as members of ODP. A representative of the Soviet Union attending the Annapolis meeting indicated that his country is seriously considering rejoining ODP, but has not yet made a decision on that score. Joseph Palca

US technology

Parallel computer unveiled

Cambridge, Massachusetts

JUST two and a half years after setting out to build the world's first "fifth generation" computer, Thinking Machines Corporation (TMC) of Cambridge, Massachusetts, last week unveiled the Connection Machine (CM), a massively parallel computer with 65,536 individual processors. TMC has already developed some remarkable applications for the new machine but the company consultant, Richard Feynman of California Institute of Technology, says the problems have yet to be dreamed up that will make full use of CM's unique abilities.

While a graduate student at Massachusetts Institute of Technology (MIT) in Cambridge, CM's spiritual father Danny Hillis came up with the blueprint for a data-level parallel processing computer. Using data-level processing leads to a more natural way of thinking about certain problems, says Hillis. For example, organizing a visual image by examining one small section at a time makes less sense than processing all the information in the image simultaneously. Data-level processing works best on problems with large amounts of data, whereas control-level parallel processing machines are most effective when the ratio of program to data is high. But control-level parallelism involves the difficult task of developing algorithms for asynchronous control of operations.

Communication among processors is the key to CM. Locally, this is done by interconnections among neighbouring processors. For general communications and dynamic reconfiguration, CM uses routers to send messages to the 64K processors arranged in a 16-dimensional hypercube at speeds of up to 3,000 megabits per second.

CM consists of a front-end computer, either a VAX or a Symbolics 3600, that sends instructions to the parallel processors. Each processor contains 4K of memory, so a fully configured CM has 32 megabytes of memory. Even with a \$3 million price tag, or \$1 million for a CM with 16K processors, prospective buyers are lining up to get hold of the new machine. TMC has so far delivered only one CM, to the Defense Advanced Research Projects Agency (DARPA), but in two months machines will be heading to MIT (2), Yale, Perkin Elmer Corporation and a second machine to DARPA.

In 1983, Hillis teamed up with Sheryl Handler to form TMC. Handler wanted to find a product that would provide a "fundamentally new way of computing". The new company quickly brought on board top scientists as consultants: Hillis's adviser Marvin Minsky and Thomas Poggio

of MIT, Stephen Wolfram of the Center for Advanced Study at Princeton and Caltech's Feynman. With \$16 million in equity capital and a \$4.7 million DARPA contract early on, TMC was able to take the risks Handler felt necessary to leapfrog the competition.

A major advantage of CM for new users is its ability to run in familiar operating environments. Extensions to C and LISP computer languages designed by TMC allow immediate access to CM's parallel capabilities. TMC has already come up with applications packages that show off CM's power. A search for key words in a 16,000-document database takes only 30 milliseconds, much faster than conventional searches. Image processing, very large scale integrated (VLSI) circuit design, and fluid flow problems all take advantage of CM by assigning one processor to each element of the problem's dataset. CM can achieve operating speeds up to 7,000 MIPS (million instructions per second) on certain applications, faster than the fastest computer utilizing more conventional architecture.

Although CM's initial capabilities are

impressive, Wolfram believes that a new computer language is needed to take full advantage of it. CM is well suited to cellular automata, and Wolfram predicts that almost all problems now being solved using differential equations will be done in future by constructs like cellular automata.

Another plus for CM is its expandability. While traditional computers using principles developed 40 years ago by von Neumann are inherently limited by a single central processing unit, an arbitrarily large number of processors can in theory be linked together using CM's design. TMC scientists have already written simulations of a machine with one million processors that can run on current CM computers.

Hillis says TMC's next goal is to develop a true learning system. In a small way, the interconnecting processors of the CM resemble the interconnections of individual neurones of the brain. Hillis hopes that this type of architecture will encourage the development of learning algorithms that more closely resemble biological processes.

Other groups around the world are working on new computers using parallel architecture, but Poggio points out that CM has one obvious advantage over other projects: "It works".

Joseph Palca

AIDS in India

Pool of infected women?

New Delhi

THE Indian Council of Medical Research (ICMR) is expanding its network for surveillance on acquired immune deficiency syndrome (AIDS) following the detection of six cases in Madras, capital of the southern state of Tamilnadu.

All those concerned are female prostitutes aged between 20 and 30. Two of them have disappeared and are being sought by the state government and the other four are under observation. Professor V. Ramalingaswami, director-general of ICMR and chairman of the AIDS task force that has been set up, says that none of the women showed clinical manifestations of AIDS except one who lost weight. The women were among 102 prostitutes whose sera were subjected to enzyme-linked immunosorbent assay (ELISA) at the ICMR centre at the Christian Medical College in Vellore.

The findings were confirmed by scientists at the Centers for Disease Control in Atlanta, Georgia, where the sera were analysed by Western blot technique.

ICMR fears that more AIDS cases may be detected once the male contacts of the prostitutes are traced, but this is an almost impossible task. The women have told social workers that they usually had ten partners a day, mostly truck drivers; they all denied having foreigners among their

customers.

It is not yet known if the virus is HTLV-III or a related virus recently found in Senegal, West Africa. As there is no experience in India of isolating or characterizing the virus, ICMR has sent three scientists to the United States and West Germany for training.

Meanwhile, ICMR is launching a programme to train physicians and technicians in the diagnosis of AIDS. Twenty-five thousand test kits have so far been imported. It is proposed to set up surveillance centres in each district, and to test donations from professional blood donors. Imported blood and blood products now have to be accompanied by certificates testifying that they are free from AIDS, and a requirement for health certificates from visiting foreigners is being seriously considered.

A separate section is being created in the health ministry to implement the AIDS detection and control programme. Ramalingaswami says this will put an extra strain on the health budget, half of which at present goes towards control of malaria. "We have our priorities and there is not much money for AIDS observation", says Ramalingaswami, but he admits that the detection of AIDS must be a priority if it is not to become widespread.

K.S. Jayaraman

Archaeology congress

US to attend or stay away?

New Orleans

ATTENDANCE at the 1986 World Archaeology Congress at Southampton, England, has become a "moral litmus test" for US archaeologists. But there was no consensus among potential participants at the meeting here last month of the Society for American Archeology (SAA).

The congress, planned over several years with special attention to the participation of scientists from developing countries, has been controversial ever since the British organizing committee sent letters of "disinvitation" to South African and Namibian scientists. In the aftermath, the international parent body, the International Union of Prehistoric and Protohistoric Sciences (IUPPS), withdrew

its recognition of the Southampton meeting; a conference to be held in Mainz, West Germany, in September 1987 is now recognized as the official 11th World Archaeology Congress (see *Nature* 320, 3; 1986).

Opinion at New Orleans seems to have been that it is for scientists individually to decide how to balance the principle of a stand against apartheid and that of free international exchange between scientists.

Thus one would-be participant said that he wanted nothing to do with a meeting that would exclude such noted scientists and opponents of apartheid as Professor Phillip Tobias (University of the Witwatersrand), but that "I don't go along

with some colleagues' wish to ostracize those who may attend".

On the other hand, Thomas Patterson (Temple University), a co-founder of Archaeologists Against Apartheid, urged colleagues to attend the Southampton congress to avoid giving the impression that US archaeologists, "consciously or unconsciously", are in favour of apartheid. According to Ian Hodder (University of Cambridge), some 22 per cent of registrants from the United States have already cancelled.

The large number of participants from developing countries expected at Southampton adds another side to the question. Hodder says that many of them would either be prevented by their governments from attending, or would abstain on their own if South Africans were to be present.

Some at the SAA meeting raised the further question of whether an international congress is possible at all. "What guarantee is there that the same problems will not recur in Mainz?", one archaeologist asked. Hodder says that anti-apartheid groups in West Germany are mobilizing to protest at that meeting.

Meanwhile, in Berkeley, California, a symposium held in mid-April to honour anthropologist J. Desmond Clark upon his retirement, with the participation of several scholars from South Africa, passed without incident, because the existence of the symposium was not made known to anti-apartheid groups there (see *Nature* 320, 562; 1986).

One project scheduled for discussion at Southampton is the establishment of a truly international body, to correct what is generally conceded to be a European bias in the present make-up of the international union. A more broadly based coalition, it is hoped, could better ensure the free assembly of its members.

James Sacket (University of California, Los Angeles) urged that the United States should bring more weight to bear in future actions of IUPPS, where its role is small although it contributes a large proportion of the world's practising archaeologists. He suggested that one step towards the greater internationalization desired by all would be to hold the 12th World Archaeology Congress, scheduled for 1991, in the United States.

A member of the IUPPS executive committee, Sacket had made several attempts to reconcile the position of the British national committee with the requirements of IUPPS, proposing for example that participants should have the option of registering simply as individuals and not in the name of their country, but was unable to effect a compromise. "I was told that 'We do not want interference from the great powers'", he said. Thus, at this point, the future role of the United States in the international union seems problematic.

Sandra Ackerman

Biotechnology

DNA fingerprints go commercial

By early next year, and for something like £250, it will be possible to obtain a DNA fingerprint from a blood or tissue sample sent to a testing centre in Britain. The initial demand is expected to be for paternity testing, but the test has already proved invaluable in proving relationships that are doubted by immigration authorities and has considerable potential in forensic testing too.

The centre is to be operated by the diagnostics group of Imperial Chemical Industries (ICI), the British chemical and pharmaceutical giant, which is in an advanced stage of negotiation with the Lister Institute for the exclusive rights to a test devised by Dr Alec Jeffries of the University of Leicester. Royalties from the tests, which are likely to be substantial, will be shared by the University of Leicester, Dr Jeffries and the Lister Institute, of which he is a research fellow.

It requires considerable expertise to carry out the test, which types individuals in terms of a set of highly variable DNA sequences on human chromosomes. At the heart of the test is a process of hybridization between the individual's DNA and radioactive DNA probes. Before the hybridization can be carried out, however, the DNA has to be extracted and cut by enzymes into fragments which are separated according to size. Even after hybridization, it is not a simple matter to interpret the "result", a series of radioisotopic bands on a gel.

Although such techniques are routine in molecular genetics laboratories, they are much too demanding for any routine diagnostic laboratory. That is why the DNA fingerprinting test will not be marketed as a kit but offered as a service. (There is one commercial kit based on

DNA hybridization which is used for the detection of salmonella, and its manufacturers, Integrated Genetics, plan others. But none of these tests will require other than crude sample preparations and they only detect the presence or absence of a DNA sequence that identifies the organism.)

Demand for the test seems certain to be considerable, with estimates that many thousands of immigration disputes worldwide could be settled each year. Dr Jeffries has been able to cope with only a few per cent of the requests that he has received in the wake of the initial publications that illustrated the potential of DNA fingerprinting (*Nature* 314, 67; 316, 76; 317, 818 and 318, 577 (1985)).

Forensic testing is likely to be more problematic because of the need for prosecution and defence to have recourse to independent tests. There is therefore considerable uneasiness about there being only one centre to which samples can be sent. One solution in Britain would be for the Home Office Forensic Science Service, which has been evaluating the test, to carry out its own DNA fingerprinting, presumably for the police, while the defence made use of the ICI centre. Another solution would be to allow separate scientists, acting for the two sides, to have independent access to a single facility, as they sometimes do now in forensic science laboratories.

Further demand for testing may come from the armed forces. In the United States, at least, there have already been discreet enquiries about the possibility of having all military personnel DNA fingerprinted to make it possible to identify from tissue fragments the victims of air disasters or bombings. **Peter Newmark**

Research planning

UK needs a forum for decision

THERE should be a mechanism in Britain for directing public investment in research towards areas likely to yield economic benefit, according to a report of the Advisory Committee on Applied Research and Development (ACARD), published this week (HMSO, £4.90). But the report does not live up to its title, *Exploitable Areas of Science*, in that it fails to pick winners in which the government might invest, concentrating instead on the considerations it believes should determine investment in research.

The report is the work of a committee under Dr C.H. Reece, the director of research at Imperial Chemical Industries, which includes both academic and industrial scientists. The starting point for the study is the British view (which the committee denies) "that research cannot be organised to deliver economic returns". One of the chief conclusions is that, in the allocation of research funds, the public research councils should be more sensitive to estimates of the kinds of research that might yield economic benefit.

Like other such committees, the ACARD group wrings its hands over Britain's economic performance in recent years. Recognizing that the flow of oil from the North Sea has been one of the influences leading to the decline of British manufacturing industry and that oil will be a declining resource in the future, the committee says that research planning now should be aimed at the development of technology that will allow the maintenance of living standards when the North Sea is less prolific.

The committee says its chief task has been to decide whether deliberate research planning is feasible; it concludes that it is. But, according to the report, there is at present no forum in which these questions can be discussed and decided. ACARD asks for a process of continuing consultation with "researchers (new and established), research allocation organisations . . . and the industrial research community".

To these ends, the committee says, there should be a more systematic approach to the collection of data on the patterns of research in Britain. This echoes a growing concern that the data available on the conduct of research in Britain, even as now gathered in the annual review published by the Cabinet Office, are incomplete, unreliable and out of date.

The committee also asks that its new mechanism for research planning should include a means by which "relevant opinions and information" can be "evaluated". The committee would rely on informed discussion to reach conclusions, but it skirts around the question of who would

be involved in such a process. Thereafter, the committee says, there should be a process for allocating funds to priority areas (at least at one point, the report uses the verb "to prioritise") and a "commitment to exploit the results of science to UK benefit".

If the government follows the committee's recommendations, there will be no change in the present arrangements for channelling funds towards research, but

more resources will be spent on the gathering and evaluation of data. The report says that "there should be a small management group" to steer the process.

The committee acknowledges that its proposals would entail "changes in emphasis" of the present work of the research councils, but says that the result should be "a more rational allocation of resources to those science areas of prime national interest". On industrial research, the committee pleads for more foresight by industrial companies with the objective of defining "potentially useful fields in which to work." □

AIDS in California

State aid offered for vaccine

Washington

LEGISLATION that would "shut the door" on strict product liability for vaccines against AIDS (acquired immune deficiency syndrome) is working its way through the California State Assembly, with the blessing of the two biotechnology companies closest to vaccine development. In addition to its shield from liability, the bill provides \$6 million in aid for clinical testing and \$20 million for state purchase of the vaccine to guarantee a million-dose market. Genentech and Chiron, eager to see the bill made law, say it would remove economic obstacles to vaccine development; California hopes it would induce these companies to triple their development efforts.

California ranks second in the United States for diagnosed cases of AIDS, with over 4,600 reported, and a study released last month by the state's Department of Health Services (DHS) predicts the disease will have struck 30,000 Californians by 1990. Feeling both the pressure of a public health threat and the promise of engineered immunization from companies virtually in the legislative backyard, assemblyman John Vasconcellos last year met with Genentech and Chiron to discuss ways to accelerate vaccine development.

The companies outlined the reasons that hormones and growth factors grab most of the attention in their laboratories, while investment in AIDS vaccine research remains modest. They said they were deterred by the uncertainty of the market size, the costs of clinical trials, and the threat of litigation once the vaccine reaches the public.

Vasconcellos and other California assemblymen drew up a bill to answer these concerns. The measure would permit the state to grant each of three companies up to \$2 million to defray the costs of clinical testing, the most expensive phase of drug development. If companies sold less than one million doses, California would buy the difference. And the state would also accept responsibility for

injuries proved to have resulted from vaccine use, with recovery limited to \$250,000 per case. The bill applies to California companies only, and they would still be held liable for cases involving negligence.

Most of California's legislators appear to favour the bill, which was introduced in January and heard in the Assembly judiciary committee last week. However, opposition may emerge next week, when the legislation goes before the Assembly's ways and means committee. But supporters can point to the DHS report, which expects medical costs for AIDS of \$3,500 million by 1990, and a California Manufacturers Association estimate that each employee who contracts AIDS costs his employer \$250,000. Faced with these figures, state-subsidized vaccines may seem like a bargain to Californians. **Karen Wright**

Degrees by consent

HIGHER degrees in Poland will in future be subject to party approval, according to a Polish scientist who emigrated last month. The PhD *Doctor Habilitatus* degrees can be conferred both by the universities and by the Academy of Sciences, but since 1982, there have been almost no formal doctoral studies in the universities (see *Nature* 320, 693; 1986), and graduates wishing to proceed to a higher degree can best do so by obtaining a research post in one of the institutes of the Academy of Sciences.

Until now, decisions on higher degrees awarded by the academy have been taken by the research council of the relevant institute. Some six weeks ago, however, Dr Zdzisław Kaczmarek, the Learned Secretary of the academy, announced that from now on, he will decide, on the basis of a confidential report on the candidate from the party organization in the institute; this report will not be accessible to the members of the research council of the institute.

Vera Rich

Future of UK observatory

SIR—Your views on the future of the Royal Greenwich Observatory (RGO) (*Nature* 320, 384; 1986) cannot be allowed to go unchallenged.

It is quite untrue that British astronomers are unperturbed by the announcement by the Science and Engineering Research Council (SERC) that RGO is to move; stunned would be a better word to describe our reaction to the extraordinary suggestion of a merger with the Royal Observatory, Edinburgh (ROE). But there is a growing ground swell of protest which will begin to surface as more people realize that they cannot this time leave action to somebody else.

I must also take issue with your use of the phrase "surplus observatory" in the heading. To put it at its lowest, if one observatory is surplus, why should it automatically be assumed that it is RGO? The telescopes at ROE are no more used for research than those at RGO. Moreover, I do not believe that either observatory is surplus to the needs of British astronomy. The two royal observatories are the major centres of research in optical astronomy in this country and it is healthy to have rivalry between them, just as it is good to have two major radio observatories and two major X-ray groups.

Each observatory runs large overseas facilities, which are complementary, and so no great benefit would accrue from administering them from the same site. A research council with vision would realize that developing the instruments for these facilities is only the first step in exploiting the overseas sites and that, far from their usefulness diminishing, both observatories can continue to play vital roles by collaborating with university astronomers to obtain the best results from the new instruments. It should not be assumed that the observatories' natural role is just to support university research; they will do that job far more effectively if they also collaborate actively in the research, as they do now.

RGO differs from ROE in not being actually on a university campus. (Actually, ROE is not quite on the campus either, but it is closer.) But that does not mean that it does not collaborate with university groups. Indeed, there has been very active collaboration with the University of Sussex for some twenty years, and we do not understand why the chairman of SERC feels that this connection is "not the kind of interaction we are looking for" (*Nature* 320, 297; 1986). There are many places where related institutions on the same campus have far less interaction. Proximity is no guarantee of interaction nor does separation preclude it.

In addition to the connection with its nearest university, RGO also collaborates

with 12 British and European university groups on astronomical instrumentation and with more than 50 groups internationally on research projects. It is hard to see how a move to a particular university would increase this collaboration. As you say in your Opinion column, the benefits of an academic partnership are indeed huge, but they are already being realized. The danger of moving RGO from its present site is that the accompanying reduction of resources will be larger than the reductions already threatened at Herstmonceux and that universities as a whole will end up losing some of the valuable resources that they currently enjoy.

In short, the diffidence that you mention in deciding what to do about RGO arises simply from the fact that there is no obvious reason to do anything. RGO is doing a good job where it is and no serious scientific case has been made for moving it. The only conceivable reason for a move would be to save money, and there is no sign that that is possible without a massive reduction in resources which would defeat the claimed object of the move.

R.C. SMITH

*Astronomy Centre, Physics Building,
University of Sussex,
Falmer, Brighton BN1 9QH, UK*

Sexist ads

SIR—In the issue of *Nature* for 3 April 1986, organic life forms were represented in your advertisements in the following proportions: four men, all wearing lab coats and engaged in "serious" scientific pursuits; one unisex doctor treating a female patient; one unisex mouse in a laboratory; one woman, dressed in tight-fitting jeans and denim jacket, exposing a portion of her lower back, smiling inanely and apparently advertising a product involved in DNA detection.

Unfortunately, the predilection of your advertisers to portray men (and in some cases animals) as serious-minded scientists, and women as dumb decorative objects, is not confined to this one issue. A notable feature of the back cover of past issues of *Nature* has been a dripping, apparently unclad blonde emerging from a lake and trying to sell air.

Please instruct your advertisers to take into account the fact that many of your readers are professional women, and are not responsive to the type of advertisement normally associated with the darker recesses of the tyre and car parts trades.

KATHLEEN WESTON, ALISON COZENS, HILLARY C.M. NELSON, LOUISE FAIRALL, LINDA AMOS, LAUREL MENGLE-GAW, FLEUR-ANGE LEFEBVRE, MELANIE J. CLARK, CYNTHIA KENYON, MARIANN BIENZ, LINDA BREEDEN,

HEINKE SCHNABEL, MARGARET ROBINSON, DONNA ALBERTSON, IVA GREENWALD, MARIA LEPTIN, DANIELA RHODES, MARTINE VERHOEYEN, LYNETTE TIGHE, CLAUDIA BEREK, SANDRA SATCHWELL, C. JANE MITCHELL, CARLOS V. CABRERA, ROB WHITE, HUGH PELHAM, SEAN MUNRO, COLIN WATTS, R.J. BAER, ANDREW J.H. SMITH, ANDREW J. NEWMAN, KIYOSHI NAGAI, ALAN FORSTER, GREG WINTER, DAN BROWN, JEFFERSON FOOTE, LUIS MARTIN, PETER SORGER, GARETH WILLIAMS, GRAHAM COOK, ANDREW FIRE, VICTOR TYBULEWICZ, STEPHAN BECK, RICHARD TURNER, HORACE DREW, TOBY J. GIBSON, TONY KOUZARIDES

*MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK*

SIR—A recent series of advertisements in *Nature* shows us women in jeans or half-naked. Is it not a bit silly for a scientific journal to feature such silent propaganda for feminist and so-called liberal attitudes? Why not give women their dignity back by presenting them in skirts?

A. BRADBURY

F. VAN STRAATEN

*MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK*

MRC unit closure

SIR—We are writing in response to the report (*Nature* 320, 478; 1986) on the proposed closure of the Medical Research Council (MRC)'s Mammalian Genome Unit in Edinburgh. We find it disturbing that the MRC secretary, Sir James Gowans, should be unable to explain the failure to find a "suitable" director for a molecular biology unit of international reputation.

The reduction in funding of MRC, imposed by the government, is no doubt the root cause of the closure. Financial uncertainty from year to year, together with the greater rewards to be attained elsewhere, produces a decline in morale that increases the rate at which scientists leave the country.

We would have expected that the closure of a molecular biology unit of high calibre would have merited rather more than a "small decision" by MRC. If it cannot find a director deemed "suitable" by its present criteria, we suggest that those criteria may need re-examining. The loss of the Mammalian Genome Unit will be yet another nail in the coffin of British science in a field which is receiving encouragement and increased financial support from more farsighted nations.

MEMBERS OF THE UNIT

*MRC Mammalian Genome Unit,
Department of Zoology,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JT, UK*

Mechanizing cellular automata

Whether or not it is possible to represent hydrodynamical systems by autonomous entities on the vertices of a lattice threatens to become an unreal issue.

THE news this week (see p. 103) of the new parallel architecture computer developed by Thinking Machines of Cambridge, Massachusetts, should be read in conjunction with two unusually contradictory papers in *Physical Review Letters* for 15 April. Both are stimulated by an as yet unpublished paper by J. Salem and Stephen Wolfram.

Wolfram is the Princeton-based physicist who carries a torch for von Neumann's model of the cellular automaton, the microscopic entity in an array whose state at any time is determined by its own state at some previous time and by those of its immediate neighbours.

Steven A. Orszag and Victor Yakhot first set out to show that cellular automata cannot be adapted in the real world to solve the problems of fluid dynamics (*ibid.*, **56**, 1691; 1986). The next paper (Margolus, N., Toffoli, T. and Vichniac, G., *ibid.*, **56**, 1694; 1986), based on the same unpublished paper, purports to reach the opposite conclusion. What is the innocent reader to conclude?

There is nothing startlingly novel about the model of the cellular automaton. Fill accessible space with microscopic cells, most simply of the same size. Then let each be occupied (or empty of) a physical entity which can be in only one of several states (most simply, finite in number), and then allow that future generations of this system will evolve, with the passage of time, according to rules applying equally to all of them (see Wolfram, S., *Nature* **311**, 419; 1984).

Such a system can obviously be a model for various aspects of the real world. If the entities occupying separate cells are members of a population, the evolution of the arrangement as time goes on may be a model of how a living population evolves. But if the cells are occupied by tiny magnets allowed to point in one direction or the opposite (for example), the system is a model for a ferromagnet.

Such systems are also capable of modelling the behaviour of ordinary dynamic systems such as those forever cropping up in physics. Anybody who doubts that has merely to recall that systems of differential equations can be approximated to by finite difference equations which may be more conveniently soluble, but which are in any case a convenient way of writing a computer program to simulate the original set of equations.

So why not think of using networks of

cellular automata as means of solving real problems in applied physics? A few simple systems are easily imagined. Suppose, for example, that the objective is to follow the time course of a plucked violin string or of a bent beam; one might approximate the system by a one-dimensional array of cellular automata whose individual states might represent the displacement of the string from a reference datum. If the evolution of the system is then described simply in terms of the instantaneous state of each element and its nearest neighbours, the approximation is certainly no worse than that familiar in the simple textbooks of applied mathematics, where only first and second derivatives seem to matter.

The argument that has now arisen stems from what must be the most ambitious goal of all, how to use networks of cellular automata to stimulate the motion of fluids with real thermodynamic properties, heat capacity, vorticity and friction, for example. This, it seems, is what Salem and Wolfram have been up to, in a paper with the title "Thermodynamics and Hydrodynamics with Cellular Automata", which it is said is "to be published".

Orszag and Yakhot are against the conclusions of that unpublished paper, Margolus and his colleagues show that the arguments are valid. A curious almost-ethical point arises at this stage. It is rare that journals publish disputatious accounts of a paper that has not yet been made accessible to readers. The issue is only nearly ethical; readers who do not care for having to read between the lines can always read some other journal.

Reading between the lines, what Salem and Wolfram have done is to argue that it should be possible to represent a real thermodynamic fluid by a system of cellular automata in the sense that the space and time averages of the properties of cellular automata will follow the behaviour of incompressible fluids which are described by the Navier-Stokes equations. Friction enters, producing heat (and robbing the system of kinetic energy).

A crucial part of the Salem and Wolfram argument (again between the lines) is that the coming of computers with parallel architecture should make these simulations more accurate and also more real.

Orszag and Yakhot do not dispute the virtues of computers made from, say, 64,000 microprocessors strung together (who could?). Their quarrel is with the

underlying physics and the computational economy of the cellular automata models. Reading between the lines, one can conclude that Salem and Wolfram represent continuous variables such as the velocity (of the fluid) in a velocity field by means of averages over groups of neighbouring cells carrying separate cellular automata. Questions then arise about the efficacy of the representation; either there are so few cells in the group over which average values represent dynamical variables in the real world that the calculation is inaccurate, or the group is large enough to be realistic but the computational work is enormously increased. The essence of the argument is that tractable networks of cellular automata will usually be unrealistic, but that realistic networks will be less easily soluble than the equations to which they are an approximation.

Margolus *et al.* write in a quite different vein. "Surprisingly enough, cellular automata can faithfully model continuum systems such as fluids . . . [but] are poorly supported by conventional scientific computers . . . more appropriate architecture can easily gain a performance factor of at least 10,000 . . ." Part of the trick (to be revealed in the unpublished paper by Salem and Wolfram) is to devise the generation rules for the automata so that microscopic reversibility is assured. Margolus and his colleagues show, among other things, some pretty diagrams in which a central blob of fluid on a two-dimensional plane spreads outwards in a growing concentric ring. The authors surprisingly boast that one simple system (a two-dimensional lattice gas) is macroscopically as well as microscopically reversible.

Quite what the dispute is about can be read only between the lines, but there is an obvious and simple resolution of it: simply build computer systems with as many microprocessors as there are atoms whose behaviour must be simulated, and make sure that they respond at least as quickly. Then the objections of the Orszags and Yakhots of this world will melt away. As it happens, Margolus and his colleagues say that it should be possible to deal with 10^{16} elements (the square of the cube root of Avogadro's number) within a decade, and to "span the gap between the microscopic and the macroscopic world" soon afterwards. That may be possible, but the question will remain of whether it is science.

John Maddox

Protein translocation

A common mechanism for different membrane systems?

from Gottfried Schatz

ALL cells transport specific subsets of their own proteins across membranes in order to segregate metabolic pathways, to communicate with other cells or to build new membranes. Because each of the different membrane types poses the same problem — how to transport a large hydrophilic protein across a hydrophobic phospholipid bilayer — it has been tempting to suppose that cells have evolved a universal solution. Until recently, however, most of the experimental evidence seemed to argue otherwise¹. Although the endoplasmic reticulum of eukaryotes seemed to translocate proteins only while they were still being synthesized (that is, by co-translational translocation), all other membrane types studied — bacterial, mitochondrial, glyoxysomal and chloroplast — can translocate completed protein chains post-translationally. Reports from several laboratories²⁻⁸ now suggest that these differences are more apparent than real, and that the mechanism of translocation may indeed be fundamentally similar, or even the same for all membrane systems.

The co-translational model of protein translocation, derived from pioneering studies on the mammalian endoplasmic reticulum⁹, is as follows. Proteins destined to be translocated into or across the endoplasmic reticulum membrane contain a specific signal sequence, usually located at the extreme amino terminus. As this sequence emerges from the large ribosomal subunit, it binds to a cytoplasmic ribonucleoprotein complex (the signal recognition particle) which stops further elongation of the polypeptide chain. Binding of the blocked ribosomal unit to a proteinaceous receptor on the cytosolic face of the endoplasmic reticulum releases the elongation arrest and causes the growing polypeptide to move across the membrane. Soon after the amino-terminal signal peptide has traversed the membrane, it is removed by a signal peptidase located in the lumen of the endoplasmic reticulum.

This model rests on several key findings^{1,9}. First, attempts to translocate completed polypeptides into microsomes (isolated vesicular fragments of endoplasmic reticulum) have always failed: translocation occurred only while the polypeptide was still being synthesized. Second, the signal sequence was removed by the luminal peptidase before the carboxy-terminal domain of the protein had been synthesized; uncleaved, full-length

precursors were thus undetectable in pulse-labelled intact cells. Finally, the signal-recognition particle arrested the elongation of secreted, but not of cytosolic proteins; the arrest of elongation could be abolished by adding mammalian microsomes, whereupon the growing secreted proteins were translocated into the microsomes. This tight coupling between synthesis and translocation seemed to be an elegant mechanism for ensuring that proteins destined to be inserted into or across the membrane were not released into the cytosol.

The co-translational model is purely operational and does not invoke an obligate mechanistic coupling between polypeptide elongation and translocation. It has, however, often been interpreted to imply that translocation is driven by chain elongation which is in turn powered by the hydrolysis of high-energy phosphate bonds. This interpretation is difficult to prove or refute: by its very nature, a co-translational mechanism makes it impossible to probe the energy requirement of translocation independently from that of protein synthesis. Moreover, it is clear that in other membrane systems there can be no such mechanistic coupling: all other membrane systems can translocate polypeptide chains post-translationally. For example, mitochondria efficiently import pre-synthesized precursor polypeptides, both *in vitro* and *in vivo*. Import across the inner membrane requires an electrochemical potential across that membrane¹⁰. Post-translational uptake of proteins is also observed with chloroplasts¹¹ and glyoxysomes¹². With chloroplasts, uptake is dependent on ATP rather than a transmembrane potential¹³. Finally, bacteria can translocate proteins across their plasma membrane post-translationally in a process requiring an electrochemical potential across that membrane^{14,15}. Although post-translational movement of proteins across bacterial membranes has long been controversial¹, it is now accepted by most workers in the field¹⁶. For a while it seemed, therefore, that nature had devised two distinct mechanisms for moving proteins across membranes, one coupled to protein synthesis and the other not.

Last year, however, obligate coupling between protein synthesis and translocation across the endoplasmic reticulum began to look less certain. For example, elongation arrest by the signal recognition particle was not detectable with all secre-

tory proteins¹⁷, or when protein-synthesizing systems from mammalian sources rather than from wheat germ were used¹⁸; this raised the possibility that elongation arrest was an artefact limited only to some secretory proteins, and only if these were synthesized in the heterologous wheat-germ system. Similarly, signal-recognition particles from which a segment of the RNA chain and/or two protein subunits had been removed no longer stopped elongation of secretory proteins, yet still mediated their translocation across the endoplasmic reticulum^{8,19}. Finally, a signal sequence engineered into an internal location could effect translocation of both flanking domains into microsomes²⁰. In this case, at least the amino-terminal domain had to be translocated after it had been synthesized. Still, there was no direct proof that a completed protein could be translocated into microsomes.

Such proof is now available. Hansen, Garcia and Walter²; Rothblatt and Meyer^{3,5}; and Waters and Blobel⁶ find that the precursor to α -factor (a secreted yeast mating hormone) can first be synthesized in a messenger RNA-dependent protein synthesizing system from yeast and then be translocated into yeast microsomes. Taken together, these experiments show that post-translational uptake into yeast microsomes can be as efficient as co-translational uptake, is insensitive to uncouplers and ionophores, and requires ATP. It will be interesting to see whether post-translational uptake into yeast microsomes can occur with proteins other than the α -factor precursor.

Energy-dependent post-translational uptake into microsomes is not unique to yeast: Mueckler and Lodish report⁴ that dog pancreas microsomes can post-translationally incorporate the human glucose transporter, a glycoprotein of relative molecular mass 54,000 containing as many as 12 membrane-spanning segments. Even though post-translational uptake in this system is at least one order of magnitude less efficient than co-translational uptake, it is increased threefold

1. Kreil, G. A. *Rev. Biochem.* **50**, 317 (1981).
2. Hansen, W., Garcia, P.D. & Walter, P. *Cell* **44**, 397 (1986).
3. Rothblatt, J.A. & Meyer, D.I. *Cell* **44**, 619 (1986).
4. Mueckler, M. & Lodish, H.F. *Cell* **44**, 629 (1986).
5. Rothblatt, J.A. & Meyer, D.I. *EMBO J.* (in the press).
6. Waters, G.M. & Blobel, G. *J. Cell Biol.* (in the press).
7. Mueckler, M. & Lodish, H.F. *Nature* (in the press).
8. Siegel, V. & Walter, P. *Nature* **320**, 81 (1986).
9. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5 (1984).
10. Schatz, G. & Butow, R.A. *Cell* **32**, 316 (1983).
11. Highfield, P.E. & Ellis, R.J. *Nature* **271**, 420 (1978).
12. Zimmermann, R. & Neupert, W. *Eur. J. Biochem.* **112**, 225 (1980).
13. Grossman, A., Bartlett, S. & Chua, N. *Nature* **285**, 625 (1980).
14. Date, T., Zwizinski, C., Ludmerer, S. & Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 827 (1980).
15. Randall, L.L. *Cell* **33**, 231 (1983).
16. Wickner, W.T. & Lodish, H.F. *Science* **230**, 400 (1985).
17. Anderson, D.J., Mostov, K.E. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7249 (1983).
18. Meyer, D.I. *EMBO J.* **4**, 2031 (1985).
19. Siegel, V. & Walter, P.J. *J. Cell Biol.* **100**, 1913 (1985).
20. Perara, E. & Lingappa, V.R. *J. Cell Biol.* **101**, 2292 (1985).
21. Ferro-Novick, S., Novick, P., Field, C. & Schekman, R. *J. Cell Biol.* **98**, 35 (1984).

when an amino-terminal fragment of the glucose transporter is synthesized off a truncated messenger RNA that encodes the first 340 amino acids of the protein and lacks the stop codon. Uptake of this amino-terminal fragment is insensitive to various ionophores, but requires cleavage of high-energy phosphate bonds⁷.

These new findings do not contradict the previous view that movement of proteins into the endoplasmic reticulum is, in fact, co-translational. But they do show that co-translational translocation is not obligatory. As this excludes a strict mechanistic coupling between elongation of a polypeptide chain and its movement across the endoplasmic reticulum, it is

now a distinct possibility that proteins move through different membranes by fundamentally the same mechanism. This mechanism is still a mystery, but efforts to unravel it will benefit from the newly developed cell-free endoplasmic reticulum translocation systems from yeast^{2,3,6}. As these systems should allow the molecular analysis of yeast mutants defective in early steps of secretion (see, for example, ref. 21), biochemists have been handed a sharp new scalpel for dissecting the machinery which moves proteins across the endoplasmic reticulum. □

Gottfried Schatz is in the Biochemistry Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

Archaeology

Late Saxon ceramic utensils

from Richard Hodges

POTSHERDS — the tableware, cooking pots and storage containers of ancient times — appear at first sight to be the least attractive component of archaeological excavations. But the study of ancient ceramics turns out to be surprisingly useful, providing detailed information about past societies (see ref. 1). The significance of this approach is well illustrated by the recent work of Alan Vince² on the mediaeval pottery of London.

The sheer bulk of potsherds unearthed in modern open-area excavations (for example, several tons from the tenth century Viking dig at York in the United Kingdom³) has resulted in a shift of emphasis in ceramic studies away from distinguishing pottery in terms of its shape and decoration and towards a system based on pottery fabrics — the clay matrix used by the potters. This fabric series can distinguish between the workshops responsible for the pottery production. The approach necessitates identifying the macroscopic features of the clay; if these features are unclear or can be assigned more precisely to a specific source, a microscopic analysis is performed. The commonest technique in this second stage of analysis is ceramic petrology, in which thin sections of the fabric are examined under a petrological microscope. In some cases the exact provenance of the clay is apparent from its mineralogical character; more often the clay has been taken from some indistinctive riverine source and only the potter's 'recipe' can be identified.

There are tons of pottery in the excavations of the Department of Urban Archaeology at the Museum of London, making a typological listing impractical. Therefore, the fabric series was classified using macroscopic and microscopic principles⁴. Each fabric embraces forms which vary according to its date of manufacture.

Potsherds discarded in the Middle Ages offer fascinating patterns of the history of the potter's craft and how and where these low-grade utensils were sold.

The Late Saxon ceramics from London illustrate this point vividly. It is highly debatable whether London was a significant community before King Alfred's time, but with the explosion of urban development in the tenth century London

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

A group of Late Saxon shelly ware vessels from various sites in London. Vessels of this type were made from a clay containing abundant fragments of the fossil *gryphaea*, characteristic of the Jurassic clays of midland England but not of the Thames Basin.

grew rapidly, so by the end of the century it was engaged in international trading. The ceramic history puts some interesting flesh on this skeleton. First, there was no Middle Saxon (pre-Alfredian) potter serving the community; the pots were made only primitively on a hand-rotated wheel. But, with what I have recently described in *News and Views* as the first English industrial revolution⁴, the role of the potter changed. Vince shows that the well-dated tenth century groups are dominated by so-called shelly wares (see figure). Petrological analysis of this fabric

shows that it contains abundant calcareous inclusions typical of clays in the Oxford region. Pit-groups dated dendrochronologically (tree-ring analysis of their timber content) to the first half of the eleventh century demonstrate a marked change in the type of pottery being discarded. The amount of shelly ware was halved and replaced by other 'imported' pots from areas such as East Anglia, and lead-glazed pitchers made in the town of Stamford also appear in eleventh century contexts in London. A distinctive local fabric, Early Mediaeval Ware, shows abundant sand-grains in thin sections and is one of various recipes used by potters operating near the city. By the end of the tenth century the products of as many as a dozen workshops had broken a virtual monopoly. Thereafter, the rise and fall of industries follows an inexorable pattern, punctuated by a succession of imported wares that reveals London's overseas connections.

The definition of fabrics has made it possible to chart the history of England's earliest craftsmen. Vince's study of London ceramics, like his earlier studies of mediaeval pottery in the Midlands, reveals the rapid development of workshops and their adaptive attitudes to widening their range of products. The scale of the industry and its fast expansion also sheds light on the customers of these potters. The large quantities of tenth century pottery from lower-ranking urban dwellings in sites like London and York, as well as from many villages, reflects a voracious demand for mass-produced commodities.

It has previously been suggested that the peasantry of Late Saxon and Mediaeval England, in common with many traditional nineteenth and twentieth century peasant societies, had minimal access to commodities⁵. In the opinion of some historians, feudal arrangements meant that rural populations had little opportunity to take advantage of urban industrial output. It is still too early to generalize definitively from these studies of mainly urban pottery assemblages, but a picture of a highly integrated, prosperous community is emerging. Potsherds provide details of industrial and commercial connections beyond the scope of the written sources. In many ways these fragments constitute a new form of historical epigraphy and to judge from Vince's results in particular, the collection and processing of archaeology's seemingly least interesting objects is well justified. □

1. Van der Leeuw, S.E. & Pritchard, A. (eds) *The Many Dimensions of Pottery* (University of Amsterdam, 1984).

2. Vince, A.G. *Medieval Archaeol.* **29**, 25 (1985).

3. Hall, R. *The Viking Dig* (Bodley Head, London, 1984).

4. Hodges, R. *Nature News and Views* **316**, 211 (1985).

5. Hilton, R. J. *Peasant Stud.* **4**, 271 (1978).

Richard Hodges is a Lecturer in the Department of Archaeology, University of Sheffield, Sheffield S10 2TN, UK.

Macroevolution

Predation by drilling gastropods

from Michael J. Benton

PREDATION is an important agent of natural selection, and it is regarded by many as an important driving force of evolution¹. There is growing evidence that the fossil record can offer unique information on the history of particular predator-prey systems, and thus complement the work of ecologists on living forms. Gastropod molluscs (limpets, winkles and other molluscs with 'coiled' shells) are major marine predators on other shelled animals, and they use various mechanical and chemical means to drill or bore through the shell (see figure). Evidence of this activity has been found in the fossil record, although until recently, the oldest precisely identified gastropod borings dated from the early Cretaceous Blackdown Sands of Southern England — about 100 million years (Myr) ago. But new evidence shows that drilling by gastropods occurred much earlier and suggests that these early gastropods had a different form from their present-day counterparts².

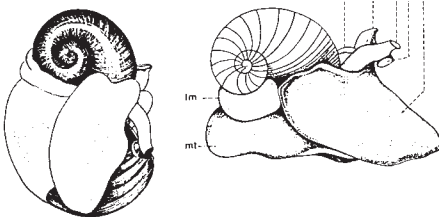
Gastropod drill holes are readily identifiable in fossils or recent specimens of shells. They are round, perpendicular to the shell surface, cylindrical or conical in cross-section, generally pass right through the shell at its thinnest point³ and are produced by the mechanical action of tooth-like structures on the gastropod radula and by the chemical action of acids.

Certain families of gastropods produce quite distinctive holes. For example, naticid gastropod drill holes are tapered inwards, forming a broad parabolic section on the outside (countersunk) and are 0.2–3.1 mm in diameter; there is a central boss in incomplete holes where the animal has stopped before reaching the flesh inside.

The fauna from the early Cretaceous Blackdown Sands of Southern England^{3,4} contains naticid gastropods and other boring forms, as well as various bivalve species. Of the 83 species of bivalves, 43 show gastropod borings, 11 of which account for 73 per cent of the attacks. One species, *Corbula elegans*, accounts for 28 per cent of the attacks. The naticids also bored into the shells of other gastropods. In general, for the Blackdown fauna, prey preference tends to vary with the relative abundance of the prey species, although some particularly thick-shelled oysters (which were quite common) were not drilled very much and some quite rare gastropod prey species were heavily attacked. The naticids selectively attacked medium-sized shells, avoiding very small ones (not enough meat inside) and very large ones (too thick to make the effort worthwhile⁵).

The siting of drill holes is also fairly constant — there seem to be preferred locations that depend on the way the predator grasps the prey shell in its fleshy foot, and on the size, shape, ornamentation and behaviour of the prey species^{4,5}.

The Blackdown Sands fauna, with its evidence of extensive naticid gastropod predation, forms part of the Mesozoic marine revolution⁶, a phase marked by the simultaneous evolution, from the early Cretaceous onwards, of powerful, relatively small, shell-destroying predators such as teleost fishes, stomatopods, decapod crustaceans and drilling gastro-



Right, *Natica josephina*. Feeding apparatus: g, 'foraging gland' secreting acid; t, tentacle; tre, respiratory tube; tro, proboscis containing radula system. pr, lm, mt comprise the hinged foot. (From Ankel, W.E. *Veh. dt. zool. Ges.* 40, 223; 1938.) Left, *N. josephina* eating a bivalve.

pods. In response, prey animals took to burrowing and evolved stronger, thicker shells that often had heavy sculpture, spines and narrow apertures^{1,4,6}. After the Cretaceous, the shell-drilling gastropods became more efficient — in the Blackdown fauna, about 4 per cent of all bivalve specimens contain drill holes, and this rose to 5–26 per cent afterwards. The numbers of drilled gastropod prey rose from 5 per cent in the Cretaceous to 11–26 per cent afterwards⁴.

There are several examples of predation by naticid gastropods before the Cretaceous. Newton⁷ reported examples from the late Triassic of North America (about 220 Myr ago) and Fürsich and Jablonski⁸ described some slightly older late Triassic cases from Italy (about 225 Myr ago). The latter bore holes show all the typical naticid characteristics, and coincide in age with the oldest known naticid gastropods. It had been assumed previously that these gastropods were initially non-drillers, and that they acquired this skill only 125 Myr later in the early Cretaceous.

Fürsich and Jablonski⁸ argue that naticid drilling ability evolved twice but became successfully established only the second time. The evidence for this assumption is the absence of naticid bore

holes in the intervening Jurassic period. In evolutionary terms, Fürsich and Jablonski view the late Triassic drillers as well enough adapted, but afflicted by bad luck — they may have had low speciation rates, or may have been wiped out in a chance extinction event. When the drilling habit was reinvented by their close relatives in the Cretaceous, it was associated with high rates of diversification, and the habit persisted to the present day in a wide range of naticid species.

But Smith and colleagues now report an even older occurrence of drilling² from the middle Devonian of North America (about 380 Myr ago). The holes are indistinguishable from those made by naticid gastropods, but they pre-date the oldest known naticid shells by 155 Myr. The bore holes were made in articulate brachiopods, and the predators show all the features of modern naticid behaviour: the drill holes are highly prey-specific in all localities sampled; they are concentrated over the 'meaty' part of the shell; and the brachiopods were attacked while still alive. As many as 44 per cent of preferred prey species were drilled, but the predators were relatively inefficient — in one locality nearly half of the drill holes were incomplete. This case forms part of the mid-Palaeozoic revolution⁹ during which several groups of predators became adapted to shell-crushing and boring (placoderm and chondrichthyan fishes, shrimp-like arthropods and possibly naticid gastropods) and the prey species evolved stronger shells with ribbing and spines in tandem.

In evolutionary terms Smith and colleagues propose that their sample documents a third origin of the naticid-like drilling habit in the Devonian, but they argue that it was not chance that terminated the first attempt at drilling. In the Devonian, there are no known naticids, nor are there any relatives that are fossilized. Smith *et al.* argue that the Devonian driller was a shell-less form (thus not fossilizable) that had evolved the physical and chemical means of drilling. Drilling is a slow process, and the predator was vulnerable itself while going about its business (hence the high number of abandoned drill holes which indicate that the predator may have been disturbed, or

1. Stanley, S.M. *Macroevolution* (Freeman, San Francisco, 1979).
2. Smith, S.A., Thayer, C.W. & Brett, C.E. *Science* **230**, 1033 (1985).
3. Bromley, R.G. *Acta Geol. Hispan.* **16**, 55 (1981).
4. Taylor, J.D., Cleavelly, R.J. & Morris, N.J. *Palaeontology* **26**, 521 (1983).
5. Kitchell, J.A., Boggs, C.H., Kitchell, J.F. & Rice, J.A. *Paleobiology* **7**, 533 (1981).
6. Vermeij, G.J. *Paleobiology* **3**, 1977).
7. Newton, C.R. *Geol. Soc. Am. Abstr. Prog.* **15**, 652 (1983).
8. Fürsich, F.T. & Jablonski, D. *Science* **224**, 78 (1984).
9. Signor, P.W., III & Brett, C.E. *Paleobiology* **10**, 229 (1984).
10. Miller, R.H. & Sundberg, R.H. *Lethaia* **17**, 185 (1984).
11. Liljedahl, L. *Lethaia* **18**, 53 (1985).
12. Stridsberg, S. *Fossils Strata* **18**, 1 (1985).
13. Sheehan, P.M. & Lesperance, P.J. *J. Paleont.* **52**, 812 (1978).

indeed eaten, before it could withdraw its drilling equipment). In either case, Smith *et al.* argue that the two 'failures' of the naticid-like shell-drilling habit in the Devonian and the Triassic were the result of poor adaptation. Only in the Cretaceous did the naticids evolve the necessary drilling techniques and the means of self-protection.

Many older examples of predatory bore holes in shells are known — for example from the late Cambrian of the United States (about 510 Myr ago)¹⁰, from the Silurian of Sweden (about 420 Myr ago)^{11,12} and from the early Devonian of Canada (about 400 Myr ago)¹³. The predators

could have been gastropods in some of these cases^{12,13} but the holes are either much smaller than those produced by living naticids^{10,11} or they lack the central boss and the inwards taper^{12,13}. The holes could have been produced by cephalopods^{3,11,12}, or by unknown forms^{3,11}. Whatever the answer, these studies offer much scope for further work on the history of marine predation, on the modelling of predator-prey interactions³ and on establishing the role of predation in macroevolution. □

Michael J. Benton is in the Department of Biology, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland.

Astrophysics

Unravelling fates of black holes

from Don N. Page

EVER since S.W. Hawking's remarkable theoretical discovery of thermal radiation emitted from black holes^{1,2}, it has been clear that these objects cannot be the stable final states of gravitational collapse they were once believed to be. But what happens during the last stages of black-hole evaporation has remained uncertain, leaving unanswered several important conceptual issues such as the fundamental predictability of the ensuing state³⁻⁵. Now M.J. Bowick, L. Smolin and L.C.R. Wijewardhana argue⁶ that superstring theories^{7,8} can help unravel the answer.

Black holes allow nothing to escape, according to classical (that is, non-quantum) physics in which each particle has a definite position and a definite velocity, which must not exceed c , the speed of light. However, in quantum physics the uncertainty principle prevents a particle from definitely being confined to the space inside the hole and simultaneously definitely having a velocity $\leq c$. In this way quantum mechanics allows a particle to tunnel out from a black hole. Hawking^{1,2} calculated this black-hole emission process in a semiclassical approximation in which quantum mechanics was applied to the particles being emitted, but not to the black hole itself, which was instead assumed to have a definite classical gravitational field.

Conservation of energy requires the black hole to lose mass, and hence shrink, as particles carry away its energy. Eventu-

ally it evaporates to such a small size that the semiclassical approximation becomes invalid, and quantum mechanics must be applied to the gravitational field of the black hole itself. It is not known how to do this in a completely consistent manner, hence the uncertainty of what happens in the last stages of black-hole evaporation. It has been variously suggested that a black hole stops radiating and shrinking when it reaches a positive-mass stable remnant; that it continues radiating until its mass becomes negative; or that it completely disappears.

There is also the question of whether the quantum evolution of the state is deterministic, or whether there is a new breakdown of predictability beyond the ordinary limitations of the uncertainty principle³⁻⁵. If the black-hole emission is precisely thermal, with no correlations between different modes (as is predicted by the semiclassical approximation), then its evaporation results in a loss of information or an increase in entropy even without any coarse graining of the final state. Furthermore, in a classical description of the gravitational field, the disappearance of a black hole would be accompanied by a so-called naked singularity (a visible edge to space-time), at which unpredictable information might enter the Universe.

Bowick, Smolin and Wijewardhana⁶ now offer a preliminary analysis of these questions in the context of superstring theories^{7,8}, exciting new candidates for a

consistent quantum 'theory of everything' (unification of the fundamental forces, including gravity; see ref. 9 for review). If one of these theories is indeed correct, it should in principle be able to explain what happens when a black hole evaporates. Unfortunately for our curiosity (but perhaps fortunately for the continuation of our careers), our understanding of superstring theories is still far too meagre to say definitely what they imply about black-hole evaporation. Bowick *et al.* readily admit this limitation, but investigate what can be said from our present knowledge of superstring theories.

By a thermodynamic analysis of perturbative (linearized) string modes, Bowick *et al.* conclude that it is statistically probable for a black hole to turn into a massive string when it gets sufficiently small. The string itself then decays into ordinary radiation (massless particles). The authors suggest that this allows the black hole to disappear completely (yet without the accompaniment of a naked singularity) because the gravitational field is only part of the superstring field.

This conclusion is eminently reasonable, although it must be admitted that the paper of Bowick *et al.*⁶ is hardly a rigorous derivation of the result. Even before superstring theories, one could have postulated that a sufficiently small black hole turns into a suitable spectrum of massive particles without having a classical gravitational field that would lead to a naked singularity. Superstring theories give hints as to how this could conceivably happen, but so far they are just hints.

On the question of whether there is a breakdown of predictability in black-hole evaporation, Bowick *et al.* admit that they do not see how superstring theory will resolve the problem. Indeed, one would need a method for correctly calculating the enormous number of multi-particle correlations between the different black-hole emission modes. Each one of these correlations is likely to be very small and hence individually close to the semiclassical approximation that it is zero, but cumulatively it is conceivable that the correlations could gradually restore to the exterior all the information that enters the hole during its formation and evaporation process. It is still too early to say whether or not superstring theory will show this possibility to be the case. □

1. Hawking, S.W. *Nature* **248**, 30 (1974).
2. Hawking, S.W. *Commun. Math. Phys.* **43**, 199 (1975).
3. Hawking, S.W. *Phys. Rev. D* **14**, 2460 (1976).
4. Page, D.N. *Phys. Rev. Lett.* **44**, 301 (1980).
5. Wald, R.M. *Phys. Rev. D* **21**, 2742 (1980).
6. Bowick, M.J., Smolin, L. & Wijewardhana, L.C.R. *Phys. Rev. Lett.* **56**, 424 (1986).
7. Green, M.B. & Schwarz, J.H. *Phys. Lett.* **B149**, 117 (1984).
8. Green, M.B. & Schwarz, J.H. *Phys. Lett.* **B151**, 21 (1985).
9. Green, M.B. *Nature* **314**, 409 (1985).

Erratum

Consensus sequence	Lys-Gly-fob-Gly-Thr-Asp-Glu-var-var-Leu-Ilu-fil-Ilu-Leu-Ala-fil-Arg
Lipocortin residues 56-72	Met-Val-Lys-Gly-Val-Asp-Glu-Ala-Thr-Ilu-Ilu-Asp-Ilu-Leu-Thr-Lys-Arg
Lipocortin residues 128-144	Lys-Gly-Leu-Gly-Thr-Asp-Glu-Asp-Thr-Leu-Ilu-Glu-Ilu-Leu-Ala-Ser-Arg
Lipocortin residues 287-303	Lys-Gly-Val-Gly-Thr-Arg-His-Lys-Ala-Leu-Ilu-Arg-Ilu-Met-Val-Ser-Arg

In the article by Robert H. Kretsinger and Carl E. Creutz "Consensus in exocytosis" (*Nature News and Views* 17 April, page 573), two of the amino-acid residues that should have been included in the figure were omitted. The correct figure is shown above.

Don N. Page is Associate Professor of Physics at The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Steroid receptors

Oncogenes as hormone receptors

from J. Michael Bishop

THE district of Hades most feared by scientists is that ignominious corner called neglect¹. This is the fate that until recently threatened *v-erb-A*, an oncogene with modest biological effect and no known function. Now however, with the discovery of structural kinship between the *verb-A* gene and the genes encoding the receptors for steroid hormones²⁻⁴, the oncogene assumes its rightful place among a family of genes whose actions help to orchestrate the specialization of cellular function that typifies metazoan organisms.

Like other retroviral oncogenes, *v-erb-A* arose by transduction of a cellular proto-oncogene (*c-erb-A*)⁵. But *v-erb-A* is unusual in that it shares the genome of the avian erythroblastosis virus (AEV) with another oncogene, the more prominent *v-erb-B*, transduced from the gene that encodes the cell surface receptor for epidermal growth factor⁶. AEV causes erythroleukaemia in chickens and transforms haemopoietic cells of the erythroid lineage in cell culture⁷. The principal responsibility for neoplastic transformation by AEV rests with *v-erb-B*^{8,9}, whose action impedes the maturation of erythroid precursor cells^{7,10}. But when *v-erb-B* acts alone, the transformed cells tend to escape the blockade and mature spontaneously⁸. The leukaemia is consequently sluggish. The cooperative action of *v-erb-A* tightens the leukaemogenic grip on the cell: spontaneous maturation becomes rare, and the leukaemia becomes correspondingly more aggressive.

What is the mechanism of this effect? Our only clues come from the structure of *v-erb-A*. The deduced amino-acid sequence of the *v-erb-A* product reveals a frail resemblance to the enzyme carbonic anhydrase¹¹. We still do not know what — if anything — that resemblance signifies; but in the face of what else has now emerged, neglect may be its fate. Our story leads instead, and unexpectedly, to steroid hormones.

The structures of steroid hormones are too simple to embody much information. The cellular receptors for the hormones must therefore serve as molecular adaptors to mount a signal of considerable complexity¹². Steroids apparently enter cells by diffusion through the plasma membrane and then combine with specific receptor proteins inside the cell. The binding of ligand 'arms' the receptor so that it in turn binds with increased affinity to select sites in cellular DNA — sites that are thought to be enhancers of transcription whose activity requires the hormone-

receptor complex. By these means a steroid hormone can induce the expression of a substantial battery of genes.

Each step in this remarkable sequence is clouded by controversy. But matters should soon become clearer because we now have molecular clones and amino-acid sequences for the glucocorticoid receptors of humans¹³, mice (G. Ringold, personal communication) and rats (K. Yamamoto, personal communication) and for the oestrogen receptors of humans^{3,4} and of chickens¹⁴. What these results make clear, first and foremost, is that all of the receptors named — and the product of *v-erb-A* for good measure — share an ancient progenitor. The structural similarities that this kinship engenders permit some powerful arguments by analogy.

There are in the receptors several domains whose conservation or other features bespeak important contributions to function. The first of these comprises approximately the amino-terminal half of the receptor molecules. Two functions have been attributed to this domain¹²: it may mediate the effect of the receptor on gene expression by interacting with the enzymatic apparatus for transcription; and it is reputed to exert allosteric influence over the remainder of the receptor. This region of the glucocorticoid receptor also commands attention because it carries prominent antigenic determinants^{12,15}, and because it contains two short stretches of very limited homology with proteins encoded by the homoeotic genes of the fruitfly *Drosophila melanogaster*². These homologies are not found in the oestrogen receptors and involve no more than six identities among amino-acid residues, so it is a tribute to the communal esteem for homoeotic genes that the homologies are given any mention at all — either here or in the original publication.

The second recognizable domain of the receptors resides near the middle of the proteins, is the most intensely conserved and harbours the site that binds DNA (refs 2-4; K. Yamamoto, personal communication). The third domain includes the carboxy ends of the receptors and extends a still-uncertain distance upstream. That this portion of the receptor binds hormone has been suspected for some time and has now been substantiated in two ways. First, two versions of the glucocorticoid receptor have been found that may arise from alternative splicing (or experimental artefact)¹³. They differ only by short substitutions of amino-acid sequence at their carboxy termini. One ver-

sion binds glucocorticoids, the other does not. The drastic effect of this small difference in structure, and the means by which similar domains of protein are nevertheless able to distinguish between glucocorticoids and oestrogens pose marvellous puzzles for the crystallographer. Second, mutations have been fashioned in the putative hormone-binding region of both the oestrogen and the glucocorticoid receptors: the fabricated mutants do not bind hormone (ref.14; K. Yamamoto, personal communication).

All of the domains within a receptor molecule may therefore have discrete functions, but they constitute a cascade, and each component may contribute to specificity in the action of steroid hormones: the ligand-binding domain is a trigger that can be thrown only by the correct hormone; the DNA-binding domain recognizes specific nucleotide sequences, whose elements of consensus are beginning to emerge¹²; and the amino-terminal half of the receptor may selectively interact with the transcriptional machinery. Cooperative division of labour of this sort within protein molecules has been elucidated in great detail with examples from bacteria and yeast.

The topographies of steroid receptors again evoke the popular scheme in which complex genes are said to have been cobbled together by genetic recombination, recruiting constituent domains from diverse parts of the evolving genome. The most adventuresome proposal is that the hormone-binding domains arose from primitive enzymes once devoted to the metabolism of steroids¹².

Whatever the origins of steroid receptors, they are apparently similar to those of the protein encoded by *v-erb-A*, which is now seen to contain domains analogous to those that in the receptors bind DNA and hormone^{2-4,14}. There is, however, no analogue in the *v-erb-A* protein for the amino-terminal half of the receptors. Instead, *v-erb-A* is fused to a portion of the viral gene *gag* that encodes virion proteins⁷. The topography of *c-erb-A*, from which the viral gene is transduced, has yet to be described. Presumably a

1. Watson, J.D. *The Double Helix* (Atheneum, New York, 1968).
2. Weinberger, C., Hollenberg, S.M., Rosenfeld, M.G. & Evans, R.M. *Nature* **318**, 670 (1985).
3. Green, S. *et al. Nature* **320**, 134 (1986).
4. Greene, G.L. *et al. Science* **231**, 1150 (1986).
5. Vennstrom, B. & Bishop, J.M. *Cell* **28**, 135 (1982).
6. Ullrich, A. *et al. Nature* **309**, 418 (1984).
7. Graf, T. & Beug, H. *Cell* **34**, 7 (1983).
8. Frykberg, F. *et al. Cell* **32**, 227 (1983).
9. Sealy, L. *et al. Virology* **130**, 155 (1983).
10. Graf, T., Ade, N. & Beug, H. *Nature* **275**, 496 (1978).
11. Debuisne, B. *et al. Science* **224**, (1984).
12. Yamamoto, K.R. *A. Rev. Genet.* **19**, 209 (1985).
13. Hollenberg, S.M. *et al. Nature* **318**, 635 (1985).
14. Krust, A. *et al. EMBO J.* (in the press).
15. Lin, P.-F. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4260 (1982).
16. King, W.J. & Greene, G.L. *Nature* **307**, 745 (1984).
17. Welshons, W.V., Lieberman, M.E. & Gorski, J. *Nature* **307**, 747 (1984).
18. Shuyser, M. & Mester, J. *Nature* **315**, 546 (1985).
19. Bishop, J.M. *Nature News and Views* **316**, 483 (1985).

region that encodes an amino-terminal component has been shorn from *c-erb-A* in the course of transduction from the gene that gave rise to *v-erb-A*. So far, there is no hint as to what the ligand for the truncated *v-erb-A* protein may be — if one indeed exists.

But the analogies we do have bring the product of *v-erb-A* into a new light: until now there was nothing to suggest that this protein might interact with nucleic acid; or that it might be activated by natural ligand; or that it might spend part of its functional life-span in the nucleus. This last point raises a question still vigorously debated by students of steroid receptors. It was once thought that the receptors in their inactive form loiter in the cytoplasm, entering the nucleus only after activation by the binding of ligand; more recent evidence places the receptors in the nucleus even in the absence of hormone^{16,17}. In either case, the subcellular location of the *v-erb-A* protein — until now thought to be the cytoplasm — deserves careful reappraisal.

What do these new findings suggest about the mechanism by which *v-erb-A* contributes to neoplastic transformation? Perhaps the simplest hypothesis is that *v-erb-A* is a transcriptional regulator run amok — a dangerously constitutive version of *c-erb-A*^{2,18}. This hypothesis ignores the possibility that the *v-erb-A* protein may lack a structural domain required for enhancement of transcription. Moreover, there are two forms of mutant glucocorticoid receptors whose properties suggest other mechanisms¹². The first of these mutant receptors binds ligand and DNA, yet fails to muster any physiological response. These 'increased nuclear transfer' or *nt*⁺ mutants (the designation reflects increased affinity of the mutant receptors for the nucleus) bear a striking resemblance to the protein encoded by *v-erb-A*: both contain only the carboxy-terminal half of the archetypal receptor molecule, including the regions that are thought to bind hormone and DNA. For the moment, we can only speculate that *v-erb-A* has structural and functional defects similar to those of the *nt*⁺ mutants. But were that speculation correct, the *v-erb-A* protein might occupy sites on DNA in a sterile manner, competing with the product of *c-erb-A* and preventing the activation of genes that would otherwise come into play during erythroid differentiation. The hypothesis demands that *c-erb-A* play a part in erythroid differentiation and that *v-erb-A* be expressed in substantial excess of *c-erb-A*; both are reasonable possibilities, but neither has yet been tested experimentally.

The second pertinent form of mutant receptor (*nt*⁻) binds glucocorticoid, but not DNA. An *nt*⁻ mutant has now been sequenced, to reveal (among other abnormalities) substitution of glycine for glu-

tamic acid at a position that might account for the defect in DNA binding (G. Ringold, personal communication). The *v-erb-A* protein has glycine at the analogous position. If the protein were similarly unable to bind DNA, it might represent an inactive 'sump' that could compete with the product of *c-erb-A* for ligand. The physiological effect would be the same as that envisioned for the *nt*⁻ mutants: a sufficient excess of the *v-erb-A* protein would nullify the action of *c-erb-A*.

I have previously argued¹⁹ that it may be fallacious to deduce the physiological functions of proto-oncogenes from the properties of their progeny oncogenes. The present instance may be an exception. If the product of *v-erb-A* does indeed act by nullifying the action of *c-erb-A*, then the genes and cells that *v-erb-A* affects are salient clues to the function of *c-erb-A* — clues that moreover deserve the attention

of anyone who is interested in normal erythropoiesis.

So once again the study of tumour viruses has conjoined with other disciplines to enrich our knowledge of cell biology, and the obscure *v-erb-A* (or more accurately, its cellular counterpart *c-erb-A*) has acquired new lustre by its accession to a gene family whose members contribute mightily to homeostasis. As one oncogene after another is equated to a previously known determinant of normal cellular phenotype, a brash question comes to mind. Might all the major devices that control the growth of vertebrate cells already be in view? Prudence forbids an answer. □

J. Michael Bishop is in the Department of Microbiology and Immunology and the Hooper Research Foundation, University of California Medical Center, San Francisco, California 94143, USA.

Animal locomotion

Three kinds of flying in animals

from R. McNeill Alexander

JEREMY Rayner and his colleagues report in this issue (*Nature* 321, 162; 1986) their work on bats that are trained to fly through clouds of helium-filled soap bubbles. These bubbles have the same density as air, so their images in multiple-flash photographs show how the air is set in motion as the bat passes through. As wings generate forces by giving momentum to the air the new results give information about the forces on the wings and allow comparison with the flight of birds.

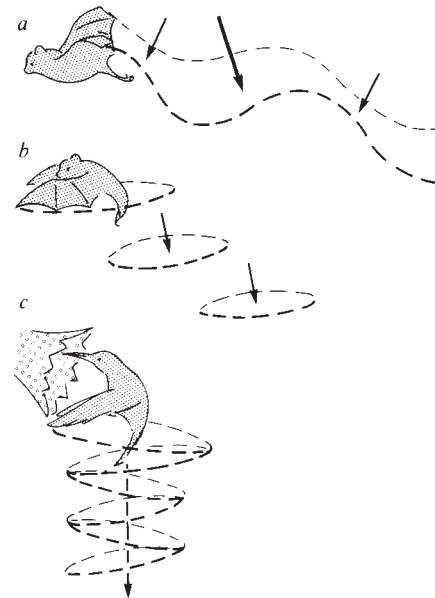


Fig. 1 Diagrams of air movements produced by *a*, kestrels and bats flying fast; *b*, pigeons and bats flying slowly and *c*, hummingbirds hovering. The broken lines represent vortices and the arrows show directions of air movements.

The experiment had been done previously with birds and gave two kinds of result (Spedding, G.R. thesis, Univ. Bristol; 1982). Pigeons produced a downward puff of air in each downstroke, and little air movement in the upstroke. Thus the downstroke produced the force that supported the animal's weight, and the upstroke was passive. Kestrels, which flew faster in the conditions of the experiment, produced downward airflow during the upstroke as well as the downstroke. Now it has been shown that Noctule bats use both techniques. In slow flight, only the downstroke is active, but at higher speeds both strokes are active. It seems likely that the two techniques are characteristic of slow and fast flight of birds and bats generally. They can be thought of as gaits, the aerial equivalents of walking and running.

Puffs or jets moving through stationary air set up vortices. Smoke rings consist of smoke trapped in the vortex around a puff of air. Vapour trails behind the wing tips of aircraft also mark vortices: the air between the two trails is moving downwards. The two flying gaits of birds and bats produce distinctive patterns of vortices, which were made visible in the experiments of Rayner *et al.* by the helium bubbles. The fast gait produces a pattern like the vapour trails of aircraft, with an important difference. The wing tips are brought closer to the body in the upstroke than in the downstroke, so the vortices undulate in and out instead of remaining parallel (Fig. 1*a*). The slow gait produces a series of vortex rings, one for each downstroke (Fig. 1*b*).

Both aircraft and flying animals need

forces to support their weight and to drive them forward through the air. Aircraft get support from their wings and thrust from their engines, but birds and bats get both from their wings, by flapping them. These mechanisms can help us to understand the difference between the gaits.

Air is driven roughly at right angles to the path of the wing through it. The downstroke drives air downwards and backwards, providing weight support and thrust. The upstroke of fast flight (Fig. 1a) drives downwards and forwards, tending to cancel out the thrust of the downstroke. To prevent this, the forces produced by the upstroke must be reduced, either by reducing the wingspan so that less air is moved by the wings, or by weakening the vortices so that the air is driven less fast. (This can be done by adjusting the angle of incidence of the wings.) Rayner has argued (*Biona Rep.* 5, 27; 1986) that it is most economical to reduce wing span and keep vortex strength (circulation) constant as kestrels and bats seem to do.

Figure 1c shows a possibility for hovering and very slow flight, when the wings can be flapped so as to move backwards (relative to the air) in the upstroke. Circulation may be reversed at the end of each stroke so that the wake is a series of linked vortex rings, a ring for each downstroke and one for each upstroke. Hummingbirds seem to do this, but the vortices have not been demonstrated. Many insects apparently are able to do the same (Ellington, C.P. *Phil. Trans. R. Soc. B305*, 1; 1984).

The slow gait (Fig. 1b) is intermediate between the other two. Circulation is eliminated but not reversed in the upstroke, so vortices (and useful forces) are produced only in the downstroke. There is scope for an optimization theory that would predict the speed at which the change should be made between the slow and the fast gaits. □

R. McNeill Alexander is Professor of Zoology at the University of Leeds, Leeds LS2 9JT, UK.

Circadian rhythms

Benzodiazepines set the clock

from A. T. Winfree

THE time will come when a traveller leaving for the airport, having frantically searched for and found his ticket, will still not be ready: it will be necessary to dash to the medicine cabinet to rummage for the inevitably misplaced bottle of jet-lag pills. Elsewhere in this issue (*Nature* 321, 167; 1986) Fred Turek and Susan Losee-Olsen, in a simple, straightforward experiment with triazolam and hamsters, provide the strongest hint yet that the elusive jet-lag pill is within reach.

Like any other mammal, humans have a persistent strong tendency to regularity in the ups and downs of diverse physiological functions, such as core temperature and the alternation of sleep and waking. The latter regularity has a native circadian periodicity close to 25 hours, and is usually synchronized by the 24-hour alternation of day and night, so it is called the biological clock. The mechanism of this clock seems rooted in the central nervous system, particularly in the suprachiasmatic nucleus of the hypothalamus. A gene whose alleles determine the periodicity of the circadian clock of the fruitfly has recently been sequenced: it appears to be a proteoglycan (Jackson, F.R. *et al. Nature* 320, 185; 1986).

Little more is known of the mechanism of circadian clocks, but their pharmacological sensitivity has recently provided us with some powerful hints about phase shifts such as those one encounters during east-west travel. Ionophores, brain hormones and neurotransmitter analogues

are conspicuous among agents capable of permanently shifting the phase of the clock, much as visible light has long been known to do. Alternating light and darkness is the leisurely agent of synchronization that normally guides plants, invertebrates and mammals (including humans) to a standard phase relationship with the environment after a shift of phase. But gradual entrainment to local time requires exposure to cycle after cycle of the synchronizer; would it not be better if we could abruptly set our internal clocks in advance, much as we set our watches? My own favoured procedure involves exposure to bright sunlight at a carefully planned hour (Daan, S. & Lewy, A.J. *Psychopharm. Bull.* 20(3), 566; 1984), but this is not necessarily convenient if, for example, it is snowing or it is night-time. Pharmaceutical assistance towards synchronization has long been sought; one need only stand in the aisle of any intercontinental jetliner to hear many superstitions about, for example, coffee, diet, alcohol and sleeping pills.

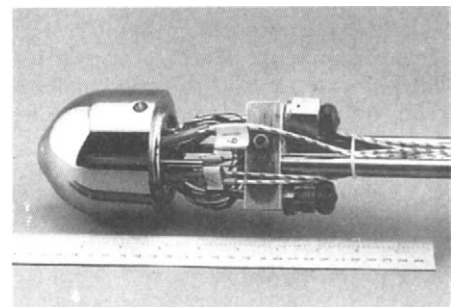
Turek and Losee-Olsen seem to have found a real effect — that of benzodiazepines — although the results of clinical trials have not yet been reported. Benzodiazepines, used for several years in the management of insomnia and related mental disorders, may act by potentiating the action of the inhibitory neurotransmitter γ -aminobutyric acid, which is known to be present in the suprachiasmatic nucleus. But its impact on circadian

rhythms has never before been tested in the standard experimental protocol used by Turek and Losee-Olsen. These authors recorded the activity rhythm of a hamster for two weeks (so that the rhythm could be reliably extrapolated to the following days), then injected the animal intraperitoneally with a dimethyl sulphoxide vehicle at any one of eight equally spaced 'times of day' in its circadian cycle. The rhythm continued as anticipated. But if the vehicle carried a benzodiazepine (2.5 mg of triazolam, equivalent to about 5 times a heavy dose for an adult human) then the timing of subsequent activity was disrupted, causing an immediate phase advance or phase delay of about one hour. The exact amount of the delay depends dramatically and quite reproducibly on the timing of the dose, a dependency unlike that caused by stimulation of the retinohypothalamic pathway for phase-shifting by light. Hence Turek and Losee-Olsen have found a new avenue of entry into the clock mechanism.

The potential practical interest of their results lies in the hope that the same or a related fast-acting, quickly eliminated chemical without conspicuous side-effects may also abruptly reset the circadian clock in humans. If so then we may soon find ourselves carrying pills and a schedule whenever we cross time zones at a pace much in excess of one per day. The schedule will prescribe when and what dose of pills to take to achieve an immediate advance or delay appropriate to the foreseen change of longitude.

The result of Turek and Losee-Olsen raises another intriguing question — in three or four of the eight recordings shown not only is there an abrupt phase shift, but also the periodicity of the clock is substantially altered for at least a week. Yet the

Forward on fusion



A REVIEW article on page 127 of this issue describes experiments that provide new optimism for the possibility of generating power from nuclear fusion. The picture above shows the target capsule used to hold the deuterium-tritium mixtures for muon-catalysed fusion experiments at the Los Alamos Meson Physics Facility in 1984. The target was electrically heated to 800 K and then cooled to 14 K with liquid helium. Up to 150 fusions per muon (average) were observed with this apparatus.

drug, if metabolized as it is in humans, should have been eliminated within several hours. Does this mean that hamsters retain much of the large dose of triazolam, or that the clock mechanism has resettable periodicity as well as resettable phase?

The experiment also raises questions about the unforeseen circadian side-effects of medications containing psychotropic compounds. David Earnest and Turek recently demonstrated (*Proc. natn. Acad.*

Sci. U.S.A. **82**, 4277; 1985) that cholinergic neurotransmission centrally mediates the effect of light on the circadian clock of the hamster. This might lead one to wonder about over-the-counter 'sleeping pills' that contain belladonna and scopolamine: do they also have an impact on circadian coordination? □

A.T. Winfree is at the Institute for Nonlinear Science at the University of California San Diego, La Jolla, California 92093, USA.

Geophysics

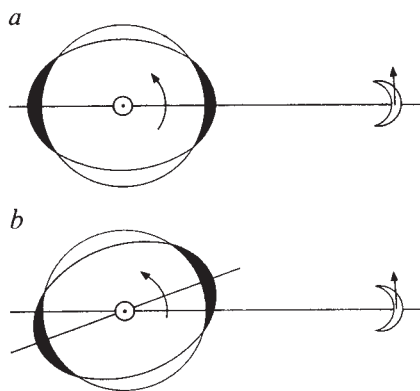
Is there coherence between Earth tides and earthquakes?

from Paul W. Burton

Do tides in the Earth cause earthquakes? Are some types of earthquakes or regions of the Earth more susceptible to earthquakes triggered by tides than others? And is this relevant to the prediction of earthquakes? There has not yet been much success in answering these questions, although the thorough studies of T.H. Heaton^{1,2} did demonstrate that some types of shallow earthquakes that have magnitudes greater than 3 are more easily triggered than are smaller events. But his later, more rigorous approach did not confirm his earlier hypothesis — much seems to be ambiguous or inextricably tangled in this subject. A. Palumbo is now trying to provide answers to some of the questions³. He finds that earthquakes most likely to be triggered by tides are those with magnitudes higher than 5 and that shallow earthquakes and their foreshocks in the Apennine Mountains are significantly more frequent when stretching in the crust caused by tides in the Earth is at a maximum.

Under consideration are stresses or tides in the solid Earth caused by the action of the Sun or the Moon on the deformable Earth. (Tides in the ocean and in the atmosphere are caused by the same actions but in more mobile and fluid media.) Stress and strain changes in the Earth caused by the varying proximity of the Sun and the Moon are accompanied by variations in the acceleration caused by gravity, g , at the surface of the Earth. Thus, g is not constant at 9.81 m s^{-2} but fluctuates both daily and over longer periods. This may be visualized simply (see figure) as deformation of the Earth's surface, and can amount to a semi-diurnal movement of the ground of as much as 0.1 m up or down. These motions are damped by friction in the Earth and so maximum strain lags behind the response expected in a totally elastic Earth; some parts of the Earth being stretched while others are compressed. Compression is maximum with the Moon or Sun on the horizon

(morning earthquakes?), and some earthquake fault systems may be susceptible to compressional perturbations. Tensional regimes across fault zones may be susceptible when the Moon or Sun are overhead (midday earthquakes?).



Earth tides caused by lunar action. *a*, Without friction, Earth strain would cause an upward motion of the ground along a line joining the centres of the Earth and the Moon. With the Moon on the horizon the strain would cause a downward and inward motion of the ground. *b*, With friction, the tidal bulge in the Earth is carried forward by rotation of the Earth causing a phase delay with respect to the action of the Moon on the Earth. These small periodic strains may perturb strain energy stored in the Earth and trigger release of that energy as an earthquake. (Adapted from ref. 6.)

When the Earth acts gravitationally on the Moon (a larger body acting on a smaller body), it is an acceptable interpretation that tidal effects trigger moonquakes. The reverse situation, when the Moon acts gravitationally on the Earth, is riddled with complexity and ambiguity: for example, the perturbations to strain are not so large; the Earth has mobile oceans and an atmosphere which varies the surface loading; and earthquake catalogues are rarely homogeneous and complete. When many globally distributed earthquakes are considered no significant effect is discernible — presumably any effect is blurred by

a multiplicity of unidentified fault orientations and stress regimes. Palumbo has deliberately isolated a relatively small region of the Earth, in Italy, and, largely rejecting Fourier spectral analysis, used the Chapman–Miller method to detect lunar daily variations (for an up-to-date account of the method with minimum mathematics, see ref. 4). Palumbo's results are clearer than his discussion and analysis: no tidal term is discernible in the Alpine or southern Calabrian areas but there is a lunar tidal term for major earthquakes and related foreshocks in the Apennines. He suggests that the tidal terms in the frequency of earthquakes are linked with extensional tidal stresses in the east–west direction.

Although hardly being on the level of prognostication associated with a 'Jupiter' effect, this suggestion is fascinating in relation to Apennine earthquakes. It will be interesting to see if the resolution of coherency achieved by Palumbo will improve when seismic moment tensor solutions (giving a full description of seismic fault orientations and slip vectors) for earthquakes in the Apennines can be combined with local tidal strain tensors.

Palumbo certainly seems to have identified one geographical region where tidal variations in the frequency of earthquakes are discernible despite the inevitable problems obscuring potential coherency between Earth tides and earthquakes. The relevance to earthquake prediction is another matter. Although much is known about Earth tides⁵ and global and local seismicity, coherency between the two is usually within the noise of any analysis; certainly earthquake disasters do not show the sort of semi-diurnal or diurnal variation which would prompt an actuary to take cover at dawn or dusk.

The problem remains largely unresolved and confused from a statistical viewpoint. But instruments used to measure Earth tides and strain changes will also be sensitive to the volume increases that are thought to accompany changes in rocks before brittle failure. Here there may be a practical link between the two sciences; otherwise coherency between tides and earthquakes appears highly localized, where it merits further study, but elsewhere, given the present state of knowledge and the resolution of this coherency, these studies appear to be rather an esoteric pursuit. □

1. Heaton, T.H. *Geophys. J. R. astr. Soc.* **43**, 307 (1975).
2. Heaton, T.H. *Bull. seism. Soc. Am.* **22**, 2181 (1983).
3. Palumbo, A. *Geophys. J. R. astr. Soc.* **84**, 93 (1986).
4. Malin, S.R.C. & Chapman, S. *Geophys. J. R. astr. Soc.* **19**, 15 (1970).
5. Melchior, P. *The Tides of the Planet Earth* (Pergamon, Oxford, 1978).
6. Kopal, Z. in *Understanding the Earth* (eds Gass, I.G., Smith, P.J. & Smith, R.C.L.) 97 (Artemis, Horsham, Sussex, 1971).

Paul W. Burton is at the British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK.

Immunology

Development and modification of the lymphocyte repertoire

from Malcolm Gelfer and Philippa Marrack

ALTHOUGH recent advances in the understanding of the genes encoding the antigen receptors of T and B lymphocytes have allowed new approaches to some of the traditional problems of immunology, they have not, as quickly became apparent at a recent meeting*, yet provided much in the way of answers. The question addressed at the meeting was that of what determines which particular T- and B-cell receptors from the extraordinarily large potential repertoire will actually be represented in the repertoire of a given animal. Some of the most lively discussion, however, (which the meeting was organized to maximize) concerned genes that are not yet known to contribute to the receptor repertoire at all.

Holes in the idiotypic net

One of the most elegant attempts to explain the mechanisms underlying the control of antibody and effector T-cell production was devised some fifteen years ago by Niels Jerne. He argued that what limits and modifies the receptor repertoire is the existence of interactions — which can be either stimulatory or inhibitory — between cells recognizing antigenic determinants called idiotypes on the antigen receptors of other cells of the immune system. Thus, for example, anti-idiotypic antibodies may be able to recognize antibodies produced by B cells and the antigen receptors of T cells, and may stimulate or suppress the proliferation of the cells that bear them. This results in an interacting network of cells and products (the idiotypic network) that influences the available repertoire before the organism is ever exposed to environmental antigen.

Although evidence in support of the theory has since accumulated, proof has been difficult to obtain, partly because the theory, although elegant and provocative, is not specific enough to allow definitive experiments to be done. With the introduction of molecular biology into immunology, it has become possible to define both the potential repertoire and the available repertoire. Moreover, much has been learned about the mechanisms whereby the receptor repertoire is expressed.

Above all, it is now known that both antibodies and T-cell receptors for antigen are encoded by genes that are assembled by somatic recombination from separate

gene segments drawn from gene pools which for some of the segments (the 'variable', or V , segments in particular) is very large and diverse. Directly or indirectly, an understanding of the molecular mechanisms by which the repertoire is generated may provide alternative explanations for some of the phenomena attributed to the operations of the network.

This applies, for example, to the recent discoveries made by Robert Kleinfeld from Martin Weigert's laboratory (University of Pennsylvania) and Michael Reth (University of Cologne). These two groups demonstrated that the genes encoding the variable region of the heavy chain of antibodies (the V_H gene segments) can undergo secondary intrachromosomal recombinations resulting in the expression of a new variable region, and thus a new recognition specificity, in an apparently already committed B cell. The mechanism of the recombination event is believed to be related to that of the normal mechanism of immunoglobulin-gene rearrangement, which depends on a conserved heptamer sequence; the site of the inter- V_H recombination appears to be a conserved heptamer within the coding region of the V_H genes. But to date, somatic inter- V_H recombination has only been seen in two B-cell lines, and it remains to be determined how general this observation will be.

Secondary recombination between V_H segments may help to explain an apparent paradox in B-cell repertoire expression: namely, that during fetal life it seems that the joining (J_H)-proximal V_H genes are preferentially rearranged (G. Yancopoulos, Columbia University), but by adulthood, all V_H segments appear to be relatively equally expressed. Perhaps the genes that undergo rearrangement early are substrates for further recombination events, the results of which shape the adult immunoglobulin repertoire.

This in turn might provide an explanation for the phenomenon of idiotypic dominance — in which in certain mouse strains most of the antibodies produced in response to a particular determinant carry a specific idiotypic. This phenomenon has been explained by network theorists as the consequence of amplification by anti-idiotypic interactions. For example, in the case of the T15 idiotypic associated with the response of some mouse strains to phosphoryl choline, John Kearney (University of Alabama, Birmingham) provided convincing evidence for the oper-

ation of regulatory events involving cascades of anti-idiotypic antibodies. (Unexpectedly, this could only be demonstrated when antibodies were administered *in utero*: in the classical network experiments, anti-idiotypic antibodies administered in adults will affect the repertoire.) Norman Klinman (Scripps Clinic Research Foundation) however suggested that the dominant idiotypic phenomenon might be explained just as well by preferential gene rearrangement to the V_H gene responsible for the T15 idiotypic.

It is possible that similar phenomena can also account for the provocative observations reported by Tereza Imanishi-Kari (Massachusetts Institute of Technology) on transgenic mice harbouring an immunoglobulin heavy-chain gene bearing the idiotypic associated with the response to nitrophenol (NP). B cells in such mice spontaneously express antibodies bearing the idiotypic carried by the product of the introduced gene — but these antibodies are not encoded by the introduced gene. Moreover, this pattern of idiotypic expression can be transferred to other mice by inoculation with T cells from the transgenic animals. On the face of it, this is strong support for an idiotypic network in which the major perturbation consequent on the introduction of the transgene is reflected in a skewing of the repertoire by regulatory T cells that are part of the net and have themselves been elicited by the (presumed) overabundance of the idiotypic encoded by the transgene early in ontogeny.

In the lively discussion that followed this presentation, participants at the meeting suggested that preferential V_H rearrangement might explain these results as well; clearly direct evidence on this, and further data from other transgenic mice, are needed.

In the meantime, other recent experiments suggest that antigen alone could account for the selection of a skewed repertoire. In most immune responses, the antibodies produced are very heterogeneous and diverse, and no two animals will be the same. But the response to some antigens seems to draw selectively on a very limited set of V regions which may be identified as the basis of the dominant idiotypic. This initial selection is followed by somatic mutation of the V regions. Experiments designed to analyse the responses to the ars idiotypic (Malcolm Gelfer, MIT), the NP idiotypic (Klaus Rajewsky, Cologne) and the Ox-1 idiotypic (K. Berek, University of Cambridge) suggest that selection by antigen may be sufficient explanation for their dominance, as in all cases so far the somatic mutations increase the affinity of the antibodies for antigen. The acceptance of non-Jernian explanations for idiotypic phenomena however will undoubtedly require more time.

The most elegant postulate of the net-

* The winter workshop on T- and B-repertoire development, organized by Eli Sercarz and Alex Miller, was held at Lake Tahoe on 22-25 January 1986.

work theory — that idiotypes shared between T and B cells result from a shared genetic pool of V segments for B- and T-cell receptors for antigen — has already been excluded by molecular biology. How the cloning of the receptor genes has advanced the understanding of the T-cell receptor repertoire was the other main topic of the meeting.

T-cell receptor

Unlike B cells, T cells are able to respond to antigen only when it is presented to them in association with a cell-surface protein, a product of the major histocompatibility complex (MHC) of the animal in question, a phenomenon called MHC restriction. There has been much debate concerning whether such a feat requires one receptor or two, the bulk of the evidence strongly suggesting that a single T-cell receptor, made up of an α/β heterodimer, is sufficient (see Robertson, M. *Nature News and Views* 317, 768; 1985). M. Steinmetz (Basel) presented further evidence from transfection experiments to support this point of view (Dembic, Z. *et al. Nature* 320, 232; 1986). He and his collaborators have used α - and β -chain genes from a cytotoxic T cell specific for the hapten, fluorescein (FL) in association

with the MHC product D^d, to transfect a T cell specific for a different hapten and MHC product and with this transferred the specificity for FL/D^d of the donor. This supports the earlier evidence that no clone-specific, variable molecule besides α and β contributes to the specificity of these cells.

The precise contribution of the α - and β -chains to the recognition of MHC and antigen, however, remains unclear. Several investigators reported the results of sequencing α - and β -chain genes from collections of T cells with related specificities for antigen and MHC, in the hope of discovering some correlation between the sequences and the specificities. S. Hedrick (University of California at San Diego) and A. Winoto (California Institute of Technology), for example, had used T-cell clones specific for a pigeon cytochrome C peptide, and with well-defined MHC reactivities (Fink, P.J., Matis, L.A., McElligott, D.C., Bookman, M. & Hedrick, S.M. *Nature* in the press). But although there were suggestions of some correlation between antigen specificity and a region of the β chain around the diversity (D) region (Fink, P.J. *et al.* and Winoto, A.), or the J region (A. Iwamoto, Toronto), J _{β} regions may also contact MHC (Iwamoto; K. Eichmann, Freiburg), so no general rules are yet apparent.

Gamma revisited

Such experiments still leave unresolved one of the major current T-cell mysteries; the problem of the role of γ chains. About eighteen months ago, Susumu Tonegawa and his colleagues at MIT reported the existence of a T-cell receptor-like complementary (c) DNA clone, the genes for which rearrange and are expressed only in T cells. Because the sequence and glycosylation patterns of the polypeptide predicted from the cDNA sequence did not match known α - or β -chain properties, the predicted protein has been named γ ; it has yet to be identified as an entity. Much is known from cDNA and genomic sequences, however. Gamma products, which were at first thought to be relatively invariant, are becoming more variable: T. Mak (University of Toronto) and D. Raulet (MIT) respectively, described a new constant (C) region and three new V-region genes in this complex, bringing the total number of known γ -chain genes in most mouse strains to four C regions (three of which cross-hybridize), each with a J, and six V regions (three of which cross-hybridize with each other). As in the mouse immunoglobulin λ locus, each V seems to rearrange to a particular J-C pair. Some V and J segments seem to be intrinsically non-functional.

The additional variability in the gene family implies that the product may play some part in specific recognition, and it

has been suggested that it is essential for the function of mature cytotoxic T cells, the cells with which it was originally associated by Tonegawa, because of higher messenger (m) RNA expression in these cells than in helper-T cells. This possibility seems to be excluded, however, not only by the experiment of Steinmetz's group, but also because most rearranged γ -chain genes in peripheral T cells are non-functional, often because they contain termination codons consequent on out-of-frame joining of V to J regions. E. Reilly, from H. Eisen's group at MIT (Reilly, E.B. *et al. Nature* in the press), for example, has sequenced 11 different cDNA clones from three cytotoxic T-cell clones and found all to be non-functional, and similar examples are available from other laboratories.

The more recently discovered non-cross-hybridizing C _{γ} genes and the three new non-cross-hybridizing V _{γ} genes leave open the possibility that the functional γ -chain cDNA in these peripheral T cells has yet to be sequenced. Alternatively, functionally rearranged γ -chain genes may be present at an early stage in the history of the T cell and be lost before maturity. In support of this notion, Raulet (MIT) showed that rearrangement and mRNA involving two particular γ -chain V regions gradually disappear from thymocytes and peripheral T cells as the animal ages. He suggested that these V regions lie downstream (3') of a third V region. A functional rearrangement to one of the 3' regions could be deleted during maturation by the subsequent rearrangement of an upstream V by a mechanism similar to that described above for V_H genes.

If γ chains do play a part at a particular stage in the life of T cells, when is this likely to be? The γ mRNA is expressed at high levels in early fetal thymocytes illustrated at this meeting by the work of R. Snodgrass (Basel) and D. Pardoll (National Institutes of Health, NIH). It is not clear that these high levels of mRNA imply that there will be high levels of protein, however, because γ -chain rearrangements are rare in hybridomas prepared from early fetal thymocytes (W. Born, Denver) and not universal in adult 'early' thymocytes (M. Davis, Stanford University). Moreover, much of the fetal thymocyte mRNA, as in mature T cells, turns out on sequencing to contain frame-shifts making it non-functional (Snodgrass). The cDNA may be overrepresented in non-functional transcripts, as Raulet reported that five out of six rearranged genomic DNA γ clones prepared from day 17 mouse fetal thymuses turned out to have open reading frames. The implications of all this, that γ proteins may be used only in small numbers for a short time during thymocyte development, and that the protein may down-regulate gene expression, are very intriguing in the light

100 years ago



Some attempts made last year at photographing the heavens by means of an instrument quite rudimentary having yielded good results, the director of the Paris Observatory gave orders for the construction of a special apparatus, the design of which is shown in the figure. This new instrument is composed of two telescopes in juxtaposition inclosed in a single metallic tube in the form of a parallelepiped, and separated from each other along their whole length by a narrow partition.

From *Nature* 34, 34; 13 May 1886.

of a role for the putative γ chain suggested by Tonegawa and Raulet.

These authors hypothesized that the γ protein is important in one of the two critical events in the control of T-cell repertoire which occur in the thymus: the selection of cells bearing receptors that recognize antigen preferentially in association with MHC products expressed in that thymus (the phenomenon of self-restriction); and in the imposition on thymocytes of tolerance to self-MHC proteins themselves. Raulet and Tonegawa propose that immature thymocytes express receptors composed of β - and γ -chains, and that those recognizing self-MHC on thymus cells are selected. Later the γ chain is replaced by the α chain of mature T cells and the affinity of the receptor for self MHC thereby eliminated except in the presence of antigen.

The feeling at the meeting was that this theory is rather unlikely to be correct, at least in its baldest form. For example, there is no evidence in the structure/function experiments described by Hedrick, Winoto, Iwamoto, Eichmann and others (see above) that a particular β chain would predictably impose specificity for a particular MHC product on receptors in which it was first combined with one chain, γ , and then with a completely unrelated polypeptide, α . But although the theory is probably not completely correct, the data dealing with expression are

becoming increasingly intriguing, and they do indicate the direction which future investigations will probably take.

Events in the thymus

Further elucidation of the events of thymic maturation will depend on the definition of the pathways of thymocyte differentiation, and these were therefore the focus of considerable interest at the meeting. The thymus contains several cell populations that have been well defined phenotypically but whose exact lineage relationships are only now becoming clear. A minimal outline of these populations and their relationships is shown in the figure.

There was general agreement that thymocytes become tolerant to MHC antigens expressed by bone marrow-derived cells (probably dendritic cells) and perhaps also epithelial cells in the thymus (R. Jordan, Newcastle University; E. Jenkinson, Birmingham University; J. Sprent, Scripps Clinic) and that they learn to recognize antigen in association with MHC molecules expressed by thymic epithelial cells (Sprent), although the experiments of D. Longo (NIH) stand in direct contradiction to this. If self-restriction is controlled by cortical epithelial cells, then receptors on cortical thymocytes must be involved in the process. One should therefore be able to learn something about this process, or perhaps tolerance induction

(although the site where this occurs is less well understood), by examining receptors on cortical cells.

Complexes of cortical thymocytes and other thymus cells have been isolated by several investigators. For example, a recent paper by A. Farr (University of Washington) (Farr, A. *et al. Cell* 43, 543; 1985) demonstrated by electron microscopy that cortical thymocytes bind epithelial cells, and that the α/β receptor is 'capped' on that part of the thymocyte surface which is in contact with the epithelial cells. This suggests that the receptor is binding some ligand, perhaps an MHC molecule, on the epithelial cell and that the process of self-restriction or tolerance induction can be seen in the micrographs.

At the meeting, A. Kruisbeek (NIH) reminded us of her past experiments showing that anti-Ia antibodies blocked the appearance of mature, L3T4⁺, class II-restricted thymocytes, but had no effect on immature 'double-positive' cells, or on Lyt2⁺ mature thymocytes (see figure). In analogous experiments, Born showed that anti-receptor antibodies also have no effect on double-positive thymocytes (although a portion of them bears receptors) but block completely the appearance of mature cells bearing receptors with which the antibody could react. Taken together, these experiments suggest that an important event involving thymocyte receptors and MHC occurs when immature double-positive cells turn into mature thymocytes and T cells. These are the first indications of the stage of thymocyte development at which self-restriction or tolerance is induced.

Jonathan Sprent summarized what is perhaps the consensus view on thymocyte maturation. Precursor cells enter the thymus, proliferate and rearrange and express their T-cell receptor and receptor-like genes α , β and γ . These cells also become double-positive L3T4⁺ and Lyt2⁺. From these cells a few are selected, by interaction of their receptors or γ -chain products with MHC products on epithelial cells, to become mature, single-positive cells. At some stage, cells that react with self-MHC, perhaps on dendritic cells at the cortico-medullary junction, are eliminated and fully mature selected cells are allowed to enter the periphery.

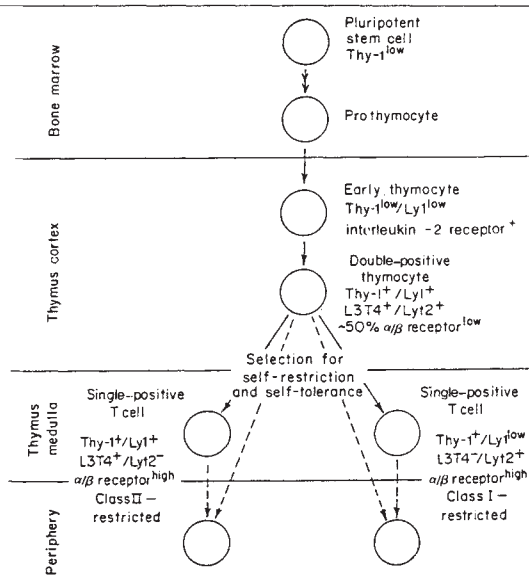
The participants left the meeting as they had come, with the feeling that many useful reagents and methods had become available for the study of lymphocyte repertoire, and with the hope and expectation that these new tools would clarify rather than confuse the situation. □

Malcolm Gelfer is in the Biology Department, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA; Philippa Marrack is in the Department of Medicine, National Jewish Hospital, Denver, Colorado 80206, USA.

Thymocyte development. Like all haematopoietic cells in adult animals, thymocytes arise from bone marrow precursors. At the meeting C. Muller-Seiburg and G. Spangrude (Stanford University) showed that the so-called primordial stem cell in bone marrow bears low levels of Thy 1. After colonization, the earliest cells found in the thymus bear little Ly1 and are positive for neither of the surface molecules that characterize mature helper or cytotoxic T cells, L3T4 or Lyt2 (T4 or T8 in man). These cells, which can repopulate an irradiated thymus, are not self-renewing, but give rise to all thymocytes sequentially, being replaced themselves eventually by new precursors from the bone marrow (B.J. Fowlkes, NIH; B. Mathieson, Frederick Cancer Research Facility; F. Jotereau, Nantes).

In elegant reconstructions of thymus organ cultures, E. Jenkinson (Birmingham) showed that a single precursor cell could give rise to all types of thymocytes. Receptors for interleukin-2 on these cells may be important in the generation of later stage thymocytes, particularly the mature L3T4⁺ or Lyt2⁺ cells (A. Kruisbeek, NIH).

Later stage cells include L3T4⁺, Lyt2⁺ blasts and non-dividing cells; and cells bearing only L3T4 or Lyt2 (like peripheral T cells). There has been considerable argument about the pathways that lead to each of these cell types. Some have suggested that the L3T4⁺, Lyt2⁺ double-positive population which comprises the bulk of thymocytes is a compulsory intermediate between the double-negative precursor and single-positive mature cell. Others believe that double-positive cells are all destined to die in the thymus. Mathieson presented at the meeting the first direct evidence for the existence of the former pathway. She selected Lyt2-bearing blast cells composed mostly of double-positive cells, and showed that after intrathymic injection, they could give rise only to cells bearing L3T4.



First isolation of HTLV-III

SIR—We wish to address several points raised in a recent commentary in *Nature*¹ regarding HTLV-III/LAV (human T-lymphotropic virus type III/lymphadenopathy-associated virus), the aetiological agent of acquired immune deficiency syndrome (AIDS). It was mentioned that the CEM cell line infected with HTLV-III/LAV may be a better source of antigen for testing HTLV-III/LAV in patients with AIDS and AIDS-related complex (ARC).

The transmission of HTLV-III isolates in several T4⁺ permanent cell lines, including CCRF-CEM, was reported first by our laboratory²⁻⁴. The transmission of HTLV-III isolates into those T4⁺ cell lines, including CEM, was the subject of a patent application made in early 1984 and now pending.

The fact that we had in our possession electron microscopic pictures of transiently transmitted LAV in Hut-78 and Ti7.4 cell lines should not be surprising. The LAV sample was obtained as a tissue culture supernatant from Dr Montagnier with the express understanding that it could be used for biomedical, biological and molecular biological studies. In accordance with this understanding, we had transiently transmitted LAV into the Hut-78 and Ti7.4 cell lines. Prior to the development of specific reagents for the detection of HTLV-III, the presence of retroviruses other than HTLV-I and HTLV-II could be detected in cultures from AIDS or ARC patients only by re-

verse transcriptase assay and electron microscopic examination.

At the time we obtained LAV it was the contention of several experts on virus morphology that the particles shown in the electron micrograph published in *Science*⁵ by Barre-Sinoussi *et al.* was an arena virus. Naturally we wanted to check the material received from Dr Montagnier by electron microscopy to check this contention. Before receipt of LAV, we had detected reverse transcriptase activity in a number of cultures from AIDS and ARC patients which showed no cross-reaction with HTLV-I or HTLV-II reagents, thus indicating the presence of a new retrovirus. In a number of cases electron microscopic examination showed the presence of virus particles with a cylindrical core, characteristic of HTLV-III (Table 1, Fig. 1). A chronological identification of some of these virus particles from cultures obtained from AIDS and ARC patients by our laboratory beginning in December 1982 is summarized in Table 1 and Fig. 1.

The experimental results shown were obtained shortly after receipt of the samples, usually a matter of weeks. We had evidence for the presence of a new retrovirus in AIDS and ARC patients long before the LAV particles were sent to us and even before the publication of the results by Barre-Sinoussi *et al.* in 1983. Since we considered the mere detection of virus particles in cultures from AIDS and ARC patients to be insufficient to confirm scientifically our hypothesis that such particles were implicated in the aetiology of the disease, we decided first to obtain specific reagents against the new virus in order to publish definitive results concerning AIDS aetiology. The results presented in our four papers provided clearcut evidence that the aetiology of AIDS and ARC was the new lymphotropic retrovirus, HTLV-III^{2,3,6,7}. In addition, for the first time the virus was produced in large

quantities, specific reagents to the virus were made, and a reliable blood test to protect the blood supply and prevent blood transfusion-associated AIDS was now available.

ROBERT C. GALLO
PREM S. SARIN
BERNARD KRAMARSKY*
ZAKI SALAHUDDIN
PHILIP MARKHAM†
MIKULAS POPOVIC

Laboratory of Tumor Cell Biology,
National Cancer Institute,
Bethesda, Maryland 20892, USA

*Electronucleonics, Inc.,
Silver Spring, Maryland 20904, USA

†Bionetics Research, Inc.,
1330-A Piccard Drive,
Rockville, Maryland 20850-4373, USA

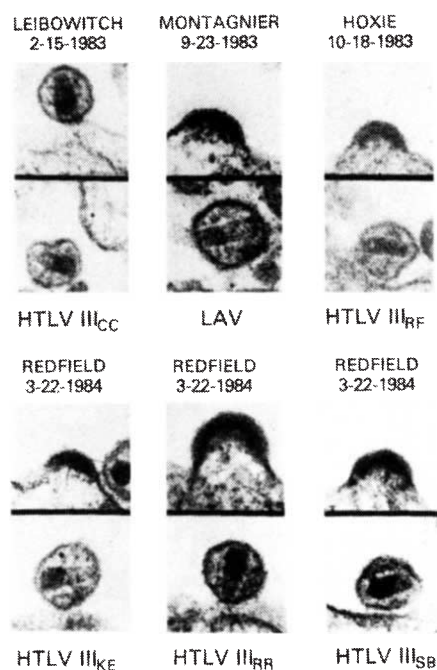


Fig. 1

Table 1 Summary of the chronological isolation of HTLV-III from patients with AIDS and ARC

No.	Patient samples	Date rec'd	Source	RT	EM	Immunological reactivities					
						HTLV-I p19	HTLV-I p24	AIDS sera	HTLV-III p19	HTLV-III p24	HTLV-III ND
1	G.W./AIDS	12/23/82	Gutterman, Houston, Texas	+	ND	—	—	ND	ND	ND	ND
2	C.C./AIDS*	2/15/83	Leibowitch, Paris	+	+	±	±	ND	ND	ND	ND
3	M.A./AIDS	2/15/83	Leibowitch, Paris	+	ND	—	—	ND	ND	ND	ND
4	B.U./AIDS	2/15/83	Leibowitch, Paris	+	ND	—	—	ND	ND	ND	ND
5	S.N./AIDS	9/23/83	Haynes, N. Carolina	+	ND	—	—	+	ND	ND	ND
6	R.F./AIDS	10/18/83	Hoxie, Philadelphia	+	+	—	—	+	ND	ND	ND
7	R.R./AIDS	2/4/84	Redfield, WRAIR, Washington, DC	+	+	—	—	+	+	+	+
8	S.S./ARC	2/4/84	Redfield, WRAIR, Washington, DC	+	+	—	—	+	+	+	+
9	K.E./ARC	2/4/84	Redfield, WRAIR, Washington, DC	+	+	—	—	+	+	+	+
10	S.B./ARC	2/4/84	Redfield, WRAIR, Washington, DC	+	+	—	—	+	+	+	+

Abbreviations used: G.W., C.C., M.A., etc., patients' initials; RT, reverse transcriptase; EM, electron microscopy; ND, not done.

*Patient C.C. was infected with both HTLV-III and HTLV-I.

1. Beardsley, T. *Nature* **320**, 563 (1986).
2. Gallo, R.C. *et al. Science* **224**, 500–503 (1984).
3. Popovic, M. *et al. Science* **224**, 497–500 (1984).
4. Popovic, M., Read-Connoles, E. & Gallo, R.C. *Lancet* **ii**, 1472–1473 (1984).
5. Barre-Sinoussi, F. *et al. Science* **220**, 868–870 (1983).
6. Shüpbach, J. *et al. Science* **224**, 503–505 (1984).
7. Sarngadharan, M.G. *et al. Science* **224**, 506–508 (1984).

SIR—I wish to take issue with a comment by Tim Beardsley on the LAV/HTLV-III controversy¹. Gilden *et al.* clearly state in their correction letter to *Science*² that I am not responsible for the erroneous mixing up of LAV and HTLV-III electron micrographs that appeared as part of a composite electron micrograph picture in my paper to *Science*³. Nevertheless, the first paragraph of Beardsley's comment evokes the impression that I am to be blamed for this embarrassing error. In fact, my work described in that *Science* paper dealt exclusively with serological and biochemical aspects of the new retrovirus. My drafts did not contain the controversial composite and neither R.V. Gilden nor M.A. Gonda were proposed as

co-authors. It was solely R.C. Gallo's decision that the composite be shifted from another paper to mine.

J. SCHÜPBACH

*Swiss Retrovirus Reference Laboratory,
Institut für Immunologie und Virologie,
Universität Zürich,
Gloriastrasse 30,
Postfach CH-8028 Zürich,
Switzerland*

1. Beardsley, T. *Nature* **320**, 563 (1986).
2. Gilden, R.V. *et al.* *Science* **232**, 307 (1986).
3. Schüpbach, J. *et al.* *Science* **224**, 503–505 (1984).

Keeping supercomputers under control

SIR—The Purdue Water Resources Research Center in conjunction with the Office of Health and Environmental Research of the US Department of Energy recently co-sponsored a meeting on the current uses of, and future needs for, supercomputers in hydrology. 'Supercomputers in Hydrology: Future Directions', the first meeting of its kind, aimed to evaluate the need for large-scale vector and parallel processing machines and their peripherals in the hydrologic sciences, especially in subsurface hydraulics. While the charge of the seminar was to specifically deal with hydrologic applications of supercomputers, the conclusions drawn at the meeting apply equally well to many other fields of science and technology.

Almost all technical issues involving the use of supercomputers were presented either by a computer scientist, a programmer or an applied mathematician. Because the architecture of most supercomputers is radically different from scalar processing machines, this is no surprise. The real importance of this observation is the implied increase in the need for multidisciplinary research teams to solve major ecological problems. Unfortunately, in the United States, many of the larger supercomputer funding initiatives do not address this issue at all.

This defect in the supercomputer initiatives leads to two major problems: scientists not doing 'science' and the inefficient use of supercomputers. If funds were available to support programmers with expertise on a machine of a certain architecture, the scientist would only need to communicate effectively with the programmer. The projected changes in architecture affect the programmer and not the scientist; hence the scientist is free to do science.

An alternative to long-run funding of programmer-consultants is for the initiatives to increase their effort in the development of very high level languages. Rice of Purdue's Computer Science Department gave a talk on projected advances in hardware and software over the next 10 years. From 1975 to 1985 computing speed increased by a factor of 25. Between 1985 and 1995 the speed is projected to increase

by a factor of 2,000. On the other hand, from 1975 to 1985, computer languages have been improved by a factor of only 1.4. Unless more funding is made available for the development of high-level languages, improvement in ease of coding is not expected to increase nearly as fast as machine computational power. In fact, because of radical changes in architecture, programming may well become more difficult.

Another important point brought out at the seminar is that advances in science should not be overpowered by advances in computational capability. Do we really improve our fundamental understanding of science by looking at larger and larger computational problems? Many would argue that the understanding of complex chemical and physical interactions in the subsurface and the ability to estimate parameters is so poor that the large supercomputer models may in fact be totally incorrect. In this instance, supercomputer modelling efforts may turn out to be a greater obstacle to real progress than no model at all.

So there is clearly a role for supercomputers, but the extent of this role will depend on the availability of high level languages, expert programmer consultants working closely with multidisciplinary teams, exposure to different computational systems, and a judicious choice of problems with the aim of advancing our understanding of science.

JOHN H. CUSHMAN

*Water Resources Research Center,
Lilly Hall of Life Sciences,
Purdue University,
West Lafayette,
Indiana 47907, USA*

FRANK J. WOBBER

*US Department of Energy,
Office of Energy Research,
OHER (ER-75),
Washington, DC 20545, USA*

Are light head loads carried free?

SIR—Geoffrey Maloiy and his colleagues (*Nature* **319**, 668; 1986) have shown conclusively that experienced African women can carry head loads at remarkably low energy cost. However, I cannot accept their conclusion that loads of less than 20 per cent of body mass are carried free, on the basis of the evidence they present. They performed no measurements in the load range 0 to 20 per cent: their conclusion is based on the backward extrapolation of a straight regression line fitted to the range of about 24 to 70 per cent of body mass. Extrapolation of a regression line always is risky, and in their case particularly so, because other data they present clearly show that the energy cost to army recruits

of carrying a back pack is not a linear function of load.

Maloiy and his colleagues suggest that an explanation for the low cost is that the women achieve a gait such that the load makes little movement against gravity during the walking cycle. Though proof of this theory using accelerometers, as R. McNeill Alexander proposes in his *News and Views* piece on the paper (*Nature* **319**, 623; 1986), would be elegant, my observations of African women carrying head loads leads me to believe that they are right. Many South African women can carry small objects like individual oranges or bottles on their heads with grace and ease; some do so by choice even when their hands are free. While it is possible to carry a metastable object like a bottle on one's head even when one's head is moving up and down, the women appear to reduce or eliminate all vertical and transverse head movements when doing so. Given the human anatomy, I think that stabilizing the head necessarily must stabilize the centre of gravity of the whole body too. Since fuelling of cyclic vertical movements of the centre of gravity constitutes one component of the energy cost of walking, elimination of such movements should reduce the energy cost. Thus, the cost of walking with a light load on the head, in people experienced in the skill, might be less than the cost of walking with no load.

We need the missing measurements.

DUNCAN MITCHELL

*Department of Physiology,
University of the Witwatersrand
Medical School,
Johannesburg 2001,
South Africa*

SIR—As supporting evidence for Maloiy *et al.*'s suggestion (*Nature* **319**, 668; 1986) about the steadiness of a head load, I cite the observation that even small children can carry liquid loads on their heads without spilling from the container. A woman carrying a bucket of water home from the well is surely one of the commonest sights; and I would think a good substitute for the suggested experiment of attaching an accelerometer to the head load.

It is interesting that it is possible to carry a much heavier load on the head than the bearer can lift unaided. It is normal to see two people lifting a burden and settling it comfortably on the head of a third, who then walks off with it. One of the most impressive, if macabre, sights I ever saw was when a farmer had died in the field, and another carried the corpse, by then rigid, to the village on his head.

D.A.H. TAYLOR

*Department of Chemistry and Applied
Chemistry,
University of Natal,
Durban 4001,
South Africa*

State of preservation of the Dead Sea Scrolls

SIR—The article "Dead Sea Scroll parchments: unfolding of the collagen molecules and racemization of aspartic acid"¹ is of interest since any method which will improve on our knowledge of vellum fragments is important. The study on the condition and deterioration of vellum which is known to be approximately 2,000 years old is interesting in itself, although the identification into goat, calf or sheep as to origin of the material examined^{2,3} would improve the information resulting from their tests. This could be accomplished by a comparison made from samples of the scrolls known to be of goat skin and that of modern goat skin prepared as we believe the scrolls to have been prepared³. In this way Weiner *et al.* could make a column to their Table 1 identifying the skin of each scroll as well as the patches and other additions. As the authors state in their paper¹ "Using different skins in one scroll and/or various parchment preparation treatments could conceivably influence the C:G index and the D/L aspartic ratios".

If, however, any relationship is to be drawn to the age of the samples and the agents of deterioration, then the tests they carried out gave insufficient evidence, as we have seen even within the scrolls there is variation. There must be a more comprehensive sampling to form a basis for comparison, which Weiner *et al.* acknowledge. A modern rat tail and a seventeenth century book cover do not provide a varied series, though the authors use them as comparative materials to the scroll samples. An adequate sample basis must be developed on a wider range of material both ancient and modern; this could be accomplished by application of the tests on samples which have a known documentation between the first and sixteenth centuries AD which would include changes in parchment preparation both in Europe and the Near East^{3,4}.

When considering the condition of materials found in the Qumran caves, it may be of interest that most scholars, given the methods of examination at the time, believed the two scrolls offered to the British Museum in the 1880's by M.W. Schapira to be forgeries based partly on the fact that their condition was throughout so superlative compared to known fragments⁵. There is now little doubt that the Schapira Scrolls came from the Dead Sea caves⁶.

It is unfortunate that Weiner *et al.* did not examine the agents of deterioration of vellum to gelatin by controlled testing, and then subject their samples to the collagen/gelatin and D/L testing. Such testing of modern and ancient samples (combined with accelerated aging tests) would

give information of a positive and precise nature on the causes of the variations found on Dead Sea Scroll samples (some on the same fragments) which undermine their study. As the authors note, we know damage has occurred, but its cause, and when it occurred are unknown. Plenderleith⁷ found during the unrolling of Dead Sea Scroll fragments of cave 1 that certain fragments were permeated with a black bituminous substance which he identified as a decomposition product of the skin. During the unrolling he found it necessary to dampen the Scrolls which caused this substance to become sticky. This black substance apparently was gelatin. This incident brings up a very important topic, that of the interventions made necessary by the fragile and rolled nature of the scrolls. Careful identification of the location from which each sample was taken should have been presented by Weiner *et al.* This information could have been correlated with published data on the condition of the fragments as they were found and treated. Both Plenderleith⁷ and Cross⁸ report extensive treatments with moisture and oils to remove clay and/or relax the skins⁹. The materials used in the treatments and the originally adherent substance could have altered analytic results.

This consideration of the condition of each fragment and their individual histories, including methods of skin preparation of which Stambolov¹⁰ has prepared a thorough survey and Poole and Reed³ a precise analysis, is of great importance. Some of the skins were found in the same cave, some scattered on the surface, others buried and some appear to have been discovered several centuries ago and been exposed to quite varied climatic conditions¹¹⁻¹³.

Others were preserved for longer periods as whole scrolls rolled in linen and placed in jars encased in tar. Varied histories of the fragments should be listed and referred to when discussing the results of testing. Rapid deterioration can occur in organic material transferred from one environment to another. Fragments taken from the dry conditions in the caves to a humid environment or one with rapid or frequent variation could cause severe and immediate damage. It is of interest to note that the authors are continuing their research with specific attention to the climatic effects of the transition of collagen to gelatin (S. Weiner, personal communication).

Given the problems with amino acid dating, especially the variation with regard to radiometric methods¹⁵⁻¹⁷, examination of rates of decay from amino acid D/L values should be considered very relative, especially with regard to natural and human interventions. Nevertheless, its use as a source of approximate time

during which collagen has changed to gelatin may prove to be of great value.

T.B. KAHLE

Capricornus School of Book Binding and Restoration,
Berkeley,
California 94701, USA

NICCOLO CALDARARO

Fine Arts Museums of San Francisco,
and *Tiburon Archaeological*
Research Group,
1600 Holloway Avenue,
San Francisco State University,
San Francisco, California 94132, USA

1. Weiner, S., Kustanovich, Z., Gil-Av, E. & Traub, W. *Nature* **287**, 820-823 (1980).
2. Ryder, M.L. *Nature* **182**, 781-783 (1958).
3. Poole, J.B. & Reed, R. *Technol. Cult.* **3**, 1-26 (1962).
4. Burton, D., Poole, J.B. & Reed, R. *Nature* **184**, 353-534 (1958).
5. Allegro, J.M. *The Shapira Affair* (Doubleday, Garden City, 1965).
6. Reed, R. *Ancient Skins, Parchments and Leathers* (Seminar Press, London, 1972).
7. Plenderleith, H.J. *Discoveries in the Judean Desert I, Qumran Cave Vol 1* (eds Barthelemy, D. & Millk, J.T.) 39-40 (1955).
8. Cross, F.M. Jr *Biblical Archaeol.* **17** (1954).
9. Araldite as an aid to Archaeology (Ciba Tech. Note No. 218, 1961).
10. Stambolov, T. *Manufacture, Deterioration and Preservation of Leather. A Literature survey of theoretical aspects and ancient techniques* (Central Research Laboratory ICOM, Amsterdam, 1969).
11. Kahle, P.E. *The Cairo Geniza* 2nd ed (Blackwell, Oxford, 1959).
12. Yadin, Y. (ed.) *The Temple Scroll* Vols. 1-3, (University of Jerusalem, 1984).
13. Vermes, G. *Times Lit. Suppl.* **3 May**, 50, 501 (1985).
14. Protsch, R.R. *Catalogue of Fossil Hominids of North America* (Fischer, New York, 1978).
15. Hedges, R.E.M. & Wallace, C.J.A. *J. Archaeol. Sci.* **5**, 377-388 (1978).
16. La Joie, K.R., Peterson, E. & Gerow, B.A. in *Biochemistry of Amino Acids* (ed. Hare, P.E.) 477-489 (Wiley, New York, 1980).
17. Kessels, H.J. & Dungworth, G. in *Biochemistry of Amino Acids* (ed. Hare, P.E.) 527-541 (Wiley, New York, 1980).

WEINER *et al.* reply—In our study, we reported the development of a new method for assessing the extent to which collagen has transformed to gelatin in ancient parchments and we tried to estimate the approximate time that elapsed since the transformation took place. We applied these methods to the analysis of some 40 parchment fragments from the Dead Sea Scrolls. We agree with Kahle and Caldaro that a thorough study, which includes more background information on the pieces analysed as well as controlled studies of parchments of different ages, prepared in various ways and subjected to a range of temperature and humidity conditions, would be the most helpful. Incidentally, some of the background information can be ascertained from the literature using the information supplied in our Table 1. We took care to emphasize in our original publication that the transformation of collagen to gelatin may be only one of several different degradative processes that occur in ancient parchments. Thus caution must be exercised when correlations between some parchment property and collagen:gelatin proportions and/or D/L ratios are made, as Kahle and Caldaro are suggesting. We also noted that the

wide range of D/L aspartic acid ratios, even from a single parchment fragment probably precludes the use of this technique for absolute age determination of parchment.

STEPHEN WEINER*
ZINA KUSTANOVICH†
EMANUEL GIL-AV‡
WOLFIE TRAUB‡

* Isotope, † Organic Chemistry and
‡ Structural Chemistry Departments,
Weizmann Institute of Science,
Rehovot, Israel

Nuclear winter and the greenhouse effect

SIR—Ever since its original publication¹, the nuclear winter concept has been in dire need of some form of experimental verification. So great has its need for independent corroboration been, in fact, that when challenged on this point^{2,3}, the concept's creators responded by grasping at some very strange straws. For one thing, they stated that the climate model they used to develop the concept was only partly calibrated, and that this calibration was accomplished by means of the possible collision of an asteroid or cometary nucleus with Earth at the time of the Cretaceous/Tertiary transition and by research on martian dust storms⁴. To my mind, a much more timely and down-to-earth phenomenon to use as a means of calibration is the CO₂ greenhouse effect; for it pertains to our own planet, is supposed to be happening now, and is itself predicted by the same climate models that produced the nuclear winter concept.

What are the fruits of such a calibration exercise? From data and equations in the US National Research Council's most recent review of the CO₂/climate connection⁵, it can be calculated⁶ that between 1880 and 1980 the northern third of the Earth (the most climatically sensitive region) should have warmed by about 3°C due to atmospheric CO₂ increases experienced over that period. However, the true model-predicted warming over this time span is actually 6°C; for an important study of the past year⁷ has clearly demonstrated that increases in other trace gases over the past 100 years should have had just as great an effect on the Earth's climate as did increases in CO₂ during that time.

So how does this 6°C warming prediction compare with reality? Recent studies of Arctic and Northern Hemisphere temperature trends of the past century^{8,10} show that the prediction is fully an order-of-magnitude too large, so it is logical to assume that the nuclear winter prediction is similarly overestimated.

Additional support for this conclusion comes from several climate modelling studies of the past two years, which have shown that when more realistic parameterizations of several real-world pro-

cesses are included in the models, the climate sensitivity of the Earth is greatly reduced. Somerville and Remer¹¹, for instance, have introduced a cloud liquid water content effect which halves climate sensitivity; while Spelman and Manabe¹² and Wang *et al.*¹³ have improved upon the treatment of feedbacks from meridional and vertical dynamical heat fluxes, including ocean currents, reducing it by a factor of four. Finally, Ou and Liou¹⁴ have improved on the treatment of cumulus convection, reducing the climate sensitivity by another 40 per cent. Taken together, these independent model improvements lead to the same order-of-magnitude reduction in the strength of the CO₂ greenhouse effect that is suggested by the temperature history of the Earth.

Why do we not see reference to these developments in the public dialogue over the CO₂/climate question? Is it because their acknowledgement would cast doubt on the nuclear winter hypothesis? These are questions that all who are truly concerned about mankind's future on our fragile planet must ask themselves. As Maddox³ has rightly said with respect to the effort to avert nuclear war, "by clouding the case with disputable predictions, they [the nuclear winter protagonists] are in danger of weakening it".

S.B. IDSO

US Water Conservation Laboratory,
4331 E. Broadway,
Phoenix, Arizona 85040, USA

1. Turco, R.P., Toon, O.B., Ackerman, T.P., Pollack, J.B. & Sagan, C. *Science* **222**, 1283–1292 (1983).
2. Maddox, J. *Nature* **307**, 107 (1984).
3. Maddox, J. *Nature* **308**, 11 (1984).
4. Turco, R.P., Toon, O.B., Ackerman, T.P., Pollack, J.B. & Sagan, C. *Nature* **311**, 307–308 (1984).
5. Carbon Dioxide Assessment Committee, US National Research Council *Changing Climate* (National Academy Press, Washington, DC, 1983).
6. Idso, S.B. *J. environ. Sci.* **27** (3), 19–22 (1984).
7. Ramanathan, V., Cicerone, R.J., Singh, H.B. & Kiehl, J.T. *J. geophys. Res.* **90**, 5547–5566 (1985).
8. Jones, P.D., Wigley, T.M.L. & Kelly, P.M. *Mon. Weath. Rev.* **110**, 59–70 (1982).
9. Kelly, P.M., Jones, P.D., Sear, C.B., Cherry, B.S.G. & Tavakol, R.K. *Mon. Weath. Rev.* **110**, 71–83 (1982).
10. Raper, S.C.B. *et al.* *Nature* **306**, 458–459 (1983).
11. Somerville, R.C.J. & Remer, L.A. *J. geophys. Res.* **89**, 9668–9672 (1984).
12. Spelman, M.J. & Manabe, S. *J. geophys. Res.* **89**, 571–586 (1984).
13. Wang, W.-C., Molnar, G., Mitchell, T.P. & Stone, P.H. *J. geophys. Res.* **89**, 4699–4711 (1984).
14. Ou, S.S. & Liou, K.-N. *J. geophys. Res.* **90**, 2223–2232 (1985).

Competition amongst desert perennials

SIR—In his *News and Views* article on regularity in the desert shrub *Larrea tridentata*, Silvertown¹ ignores some areas of current debate and seems to draw spurious analogies between intra-specific competition in plants and inter-specific competition in animals. Intra-specific interactions amongst animals may sometimes create regular spacing^{2,3}.

King and Woodell⁴ explicitly stated how the competition hypothesis could account

for regularity in desert perennials. The elimination of small individuals by root competition with neighbours could convert a clumped population first to a random and then to a regular pattern. Silvertown not only ignores this hypothesis but also states that no hypothesis was provided to explain regularly-spaced clumps. King and Woodell⁴ added that some regular patterns may have originated during long droughts, when intra-specific competition for water would be at its most intense. This is not 'acknowledging the importance of other factors'¹ as drought will inevitably intensify competition for water.

Larrea tridentata populations develop over such long time spans⁵ that the hypothesis that regularity is largely due to competition for water will be difficult to disprove. Genets of *Larrea* can live for up to 11,000 years⁶, the shrub grows slowly, and intense droughts are sporadic. Thus the lack of major change in the patterns of *L. tridentata* over 45 years in an area of fairly high rainfall⁷ is unremarkable and provides no evidence for or against any of the hypotheses which suggest how regular patterns might arise.

The effects of factors other than competition for water on pattern in *Larrea*, such as allelopathy⁸ and run-off⁹, have been discussed for a long time. Nevertheless, there is regularity in single-species populations of other dryland perennials¹⁰. It has been shown that adjacent shrubs of *L. tridentata* compete for water^{11,12}. Any refinement of the hypothesis must include (1) the clonal growth pattern of *L. tridentata* in certain sites, (2) the possibility that seedlings become preferentially established on mounds remaining after clumps have died, thus perpetuating existing patterns^{13,14} and (3) protection of young plants by older plants against freezing¹⁴.

At some sites clumps which consist of mixtures of genets may have been classified as individuals. The regularity of the clumps still needs an explanation, however, whether a clump consists of one genet or several.

T.J. KING

S.R.J. WOODELL

Department of Plant Sciences,
South Parks Road,
Oxford OX1 3RB, UK

1. Silvertown, J.W. *Nature* **316**, 298 (1985).
2. King, T.J. *J. Ecol.* **65**, 235 (1977).
3. Nelson, J.B. *The Sulidae* (Oxford University Press 1978).
4. King, T.J. & Woodell, S.R.J. *J. Ecol.* **61**, 761 (1973).
5. Sternberg, L. *Madrono* **23**, 408 (1976).
6. Vasek, F.C. *Am. J. Bot.* **67**, 246 (1980).
7. Schlesinger, W. & Jones, C. *Bot. Gaz.* **145**, 116 (1984).
8. Knipe, D. & Herbel, C.H. *Ecology* **47**, 775 (1966).
9. Anderson, D.J. *J. Ecol.* **59**, 555 (1971).
10. King, T.J. & Woodell, S.R.J. *J. Ecol.* **72**, 295 (1984).
11. Fonteyn, P.J. & Mahall, B.E. *Nature* **275**, 544 (1978).
12. Fonteyn, P.J. & Mahall, B.E. *J. Ecol.* **69**, 883 (1981).
13. Wright, R.A. *Arizona Acad. Sci.* **6**, 58 (1970).
14. MacMahon, J.A. & Schimpf, P.F. in *Water as a Factor in the Biology of North American Desert Plants* (eds Evans D.D. & Thomas, J.L.) 114 (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1981).

Growth into smallness

Nicholas Kemmer

Inward Bound: Of Matter and Forces in the Physical World. By Abraham Pais.
Oxford University Press: 1986. Pp. 637. £20, \$24.95.

Inward Bound is an account of how, since the beginning of this century, physicists have progressed towards the understanding of natural phenomena in ever-decreasing dimensions of space. The author, Abraham Pais, can look back on a long and creative career in theoretical physics, towards the end of which his biography of Einstein (*Subtle is the Lord*, published by Oxford University Press in 1982) revealed that he is a remarkably gifted writer. Now he has presented us with a book in which even more varied talents are in evidence.

Pais's mastery of the whole field of elementary particle physics is manifest on every page. In addition, his insight into the personalities of the actors in the story is remarkable. In the case of the earliest of these there is evidence of deep study of the widest range of relevant literature, but it may surprise younger readers that so many persons, whom they only know as famous names from the past, were real people in Pais's life. In particular he had the good fortune to work beside Niels Bohr in Copenhagen and Albert Einstein in Princeton. The many personal facts — sometimes of real relevance, sometimes just anecdotal — are a feature of the book, which is also enlivened by occasional spurts of impish humour. (Just one example: early on Pais tells us of speculations that radioactivity might be a property of *all* matter, and writes, "... the conjecture was not heard of again until... the 1970's..., see Chapter 22, which the reader will have time to reach without fear of prior disintegration".)

The scope and level of *Inward Bound* are clearly and extensively stated in its opening chapter, entitled "Purpose and Plan". No reader should skip it. All the chapters are divided into sections, and Section (a) of Chapter 1 is headed "From X to Z"; X is the X-ray and Z is the particle discovered by CERN in 1983, the two objects that are the termini of Pais's inward-bound journey. He explains here why that journey is broken into "A History" (c. 400 pages) and "A Memoir" (c. 175 pages). With the pace and volume of physics research increasing so enormously over the near-century covered in the book, it is hardly surprising that it was necessary to change the nature of the presentation at a particular moment. No

break could be more natural than the end of the Second World War, which also happened to coincide with the start of the author's own career in the United States. Thus the first part covers all that for Pais was there at the start, while the second is what he witnessed and helped to shape.

Even with this break, it is quite astonishing how, in spite of the explosive growth of the subject, Pais has managed to



Talk of physics: (left to right) Robert Oppenheimer, Paul Dirac and Abraham Pais.

knit events into a unified flow of narrative. This is to a large part due to the way he has chosen to unfold his story. It is not chronological, but rather topic by topic. Each of these is a strand, usually a single section, in a chapter that binds several strands together. Some of these are quite short — one might be on instrumentation, another on some general topic peripheral to the main narrative (there are for instance a few pages on physicists during the First World War) — but others are parts of long threads that can be traced almost right through the whole book, the puzzles of β -decay being an outstanding example. It is as if Pais were guiding the reader's eye along the colours in the weave of a beautiful tapestry. (I am tempted to misapply the current jargon word *fibre-bundle* to what I am trying to describe!) A feature of this particular tapestry is, of course, how it widens and thickens as one looks along it; at one end it displays all the physics that then existed, while at the other all we are following is a single, narrow strip in a huge field of activity.

In illustration, a closer look at Chapter 2 may be useful. Section (a) focuses on Roentgen in his environment at Würzburg, on how his accidental discovery came about in the wake of the introduction of cathode rays by Hertz and Lenard, on how his discovery was received and on its impact on the very private and shy dis-

coverer. Section (b) then concentrates on Becquerel, or rather the Becquerel dynasty, to start with. H. Becquerel's discovery is described against this background, and with it Chapter 2 ends. However, Chapter 3 carries this second strand further, in the story of the Curies linked to that of Rutherford and his α - and β -rays. These relatively short chapters are followed by a longer fourth, entitled "The First Particle", in which the figure of J.J. Thomson naturally enters, but not at the start. The previous, accidental discoveries (with Rutherford's work taken no further than the analysis of α - and β -radiation) did not need a great deal of background. Here, when the notion of elementary particles moves from speculation into physics, Pais needs to present more prehistory. So the first strand deals with the experimental tools that Thomson (not primarily an experimenter) owed to Geissler and Rühmkorff, and the next two with the ideas that were in the air for Thomson to bring to fruition. Section (b) is entitled "Faraday, Maxwell and the Atomicity of Charge", and Section (c) "Of Rowland, Zeeman and Lorentz". Only then does Section (d) get on to Thomson by way of Wiechert and Kaufmann. This still doesn't terminate this particular strand of

narrative; Section (e), " β -rays are Electrons", ties it to what we learnt about before, and then, finally, as a link with what is to come, we have Section (f), "Relativistic Kinematics".

At this point one might well ask how a book of this size can say anything meaningful about Faraday, Maxwell, Lorentz and Einstein, all in a few pages? The answer is in part that the exposition is a model of how clarity can be combined with brevity, but also — and this is an outstanding strength of the whole book — that every section ends with a detailed set of references to books, review articles and papers, positively inviting further study.

Having described the early chapters in such detail, I must be brief with the rest. We hear of the struggle to understand radioactivity before the picture of the nuclear atom had been thought of. Then the first piece of the story of β -decay is presented, followed by a chapter entitled "Atomic Structure and Spectral Lines" where Rutherford, Bohr and the nucleus enter the scene. A short chapter then looks briefly at how the failure of classical mechanics brought confusion into the minds of physicists, and how a generation of young men emerged, poised to sort things out. Then quantum mechanics arrives. In a chapter labelled "An Essay", Pais states emphatically that here the reader should turn to good textbooks for

complete guidance, but this does not prevent him from writing a brilliant account in which the central ideas — and equations — are presented in the midst of much on the personal side.

This is the point where the split of physics into many branches began. The narrative narrows, but an enormous amount of material nevertheless remains to be covered in the closing five chapters of the "History". It could not have been done with greater artistry. The interwoven strands tell us about quantum electrodynamics and the positron, about the continuing mystery of β -decay, leading up to the neutrino, about the merging concepts of fields and of particles that are born and die, and how a deepening understanding of symmetries has proved to be the most reliable guide to progress. In all this Pauli and Dirac, among others, are most vividly portrayed.

Inevitably, much prominence has to be given to the great shadow cast on physics by the knowledge that the theoretical structures used at the time were quite unsound mathematically, and that the necessarily approximate calculations were completely unreliable. The "History" was bound to end on such an unhappy note, but Pais has waited to its last chapter to pick up a final, more optimistic theme, going back to 1932 and the discovery of the neutron. At last there was then a convincing picture of nuclear structure to develop, into which Fermi managed to inject an at least provisionally acceptable description of β -decay. It was from there that the search for an account of nuclear forces started and, to close the story, we hear how Anderson's discovery of the muon and Yukawa's theory of the pion (as they were later to be called) seemed to open new vistas. It is left to the subsequent "Memoir" to tell us how wrong our ideas at that time were. I say "our" because this last episode of the "History" is the first of what might have been "Memoir" for me. I can vouch for the complete authenticity of the story.

The "Memoir" starts with the entirely new situation in physics after the Second World War. Many techniques, many young trained minds and large resources become available, in particular in the United States. Having survived years of danger and tragedy in his native Netherlands, Pais is there to take part in the explosive growth of his subject. The "Memoir" is the story of his life in physics and with physicists, in the centre of unprecedented developments in both experiment and theory. These Pais manages to absorb and to reproduce for the reader, still vividly but inevitably in a more technical manner than many readers (including trained physicists not on the inward-bound track) will only be able to follow in detail thanks to the extensive references.

There are just four chapters in the

"Memoir" and it would be difficult to enumerate everything in them. The "strand" structure is as before, the first three chapters telling the story up to 1960. We learn of the transformation of theoretical thinking with the arrival of the idea of renormalization and its success in explaining observed facts in electrodynamics; of the impact of the big machines and of their characteristics; of the discovery of many new particles and of the new symmetries that help to put some order into the picture; of the shock when long-cherished old symmetries turn out to be violated; and of many other avenues that are explored, some fruitlessly.

How Pais manages to fit in an account of the still-more-turbulent developments from 1960 onwards into one chapter of a

mere 70 pages (with eight pages of references to end with) is almost miraculous. He shoots the discoveries and ideas at the reader with machine-gun rapidity — but it is still the same author writing, keeping the human side up with the technical.

If I who have lived through the whole period of the "Memoir", but with far less first-hand experience of the great events, ever attempt to get my bearings on these modern things, from quarks to the Z particle, with the description of things like colliders thrown in, I know I shall turn to Pais, both his narrative and his references. It is an inimitable work. □

Nicholas Kemmer, 35 Salisbury Road, Edinburgh EH16 5AA, UK, is Emeritus Professor of Mathematical Physics at the University of Edinburgh.

Membrane steps

Miles Houslay

Reconstitutions of Transporters, Receptors, and Pathological States. By Efraim Racker. Academic: 1985. Pp. 271. Hbk \$28, £24.50; pbk \$12.95, £12.

Efraim Racker is one of the founding fathers of membrane reconstitution, and the prospect of reading his book on the subject rather intrigued me. What would be his perception of the past, present and future of membrane biology? In the preface Racker says that the book is based on a lecture series he was used to giving, which sounds promising. Each of the lectures (12 in all) is dedicated to putting over a series of 34 "lessons" varying from good general advice to homespun philosophy. A certain amount of sound advice is indeed to be welcomed by the aspiring "reconstitutor", as taking membranes apart and putting them back together is like gardening, and requires the equivalent of green fingers and a certain feel for the subject.

Lectures 1 and 2 have the central theme of "how to solubilize the protein that you've chosen". Here Racker makes the important point that what works for one protein may well not succeed for another (that is, you often have "to suck it and see"), and lists a variety of parameters of which one has to be aware. He makes it abundantly clear that there is no self-evident set of rules to follow: there are too many parameters and we have too little basic appreciation of how individual membrane proteins are glued into bilayers.

However, having said all this, the reader is then offered little or no constructive help. We are advised to take account of "critical micellar concentrations" and hydrophil-lipophil balances, but without even being told what they are, never mind how such values may be of use to tease

proteins out of their membrane home. Purification of membrane proteins, which is after all the main goal of many people interested in this field, gets a desultory page without any references to different ways of coping with this key problem. A series of general methods for reconstitution then follows, with lists of some reconstituted systems and their "activities". These lists (again innocent of references) are of little use except to give a broad view of the variety of systems that the author has got hold of and stuck back into liposomes, though the accompanying, anecdotal text does convey some idea of the principles involved.

The succeeding chapters all suffer from similar drawbacks. The section headings identify important points that should be investigated or considered, but the monologue that follows is superficial and jumps haphazardly from one system to another. One can have a degree of sympathy with such an approach in the introduction, but after three chapters one finds oneself desperately hoping to come across a logical presentation of hard facts. As it is, over the next seven chapters we are led a merry dance around a variety of membrane systems. The penultimate lecture is about cancer; it raises a few interesting points about glycolysis but is largely an indulgence, as is the final lecture centring on the author's current predilections.

I am afraid to say that after having had such high hopes for this book I found it difficult to read to the end; it lacks both substance and cohesion. Further, although Racker apologizes for all those people he has not mentioned, the number of self and selective citations in places leads to history being re-written. I would not recommend the book to students, but lecturers running courses on membranes, or sociologists of science, might find it of some interest. □

Miles Houslay is Gardiner Professor in the Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK.

Illusion in ascent

Peter J. Bowler

Taking Darwin Seriously: A Naturalistic Approach to Philosophy. By Michael Ruse. Basil Blackwell: 1986. Pp.303. £22.50, \$24.95.

MICHAEL Ruse believes that scarcely anyone has taken Darwin seriously up to now. Efforts to apply evolutionism to the problems of philosophy have failed, largely because of a refusal to acknowledge the central tenet of Darwinism: the non-progressive character of natural selection. Whatever one's opinion of Darwinism (Ruse, of course, accepts it as the basis of all legitimate evolutionary thinking), the claim that the theory has seldom been taken seriously can certainly be endorsed by the historian. To depict the growth of evolution theory as a simple ascent from Darwinism to the Modern Synthesis is to fall into the worst kind of Whig history, ignoring the large number of thinkers, including scientists, who have found the principles of natural selection unacceptable. As Ruse points out, modern Darwinists from Julian Huxley to E.O. Wilson have also found it difficult to accept the non-progressionist character of the theory.

Ruse's purpose here is to address the philosophical, not the historical issues. He criticizes evolutionary approaches to the problem of knowledge, from Herbert Spencer to Karl Popper. Drawing a simple analogy between evolution and the growth of science does not work, he says, because science is progressive and evolution is not. He looks instead at the kind of reasoning that we might expect natural selection to have programmed into our ancestors so that they could deal successfully with the real world. We believe that an effect *must* follow its cause, because any of our ancestors who did not have this mental reinforcement to back up their perception of causal regularities would soon have disappeared. The intellectual ancestor of this approach is David Hume, who realized that we read necessity into nature even though we have no philosophically satisfactory guarantee that there is anything corresponding to that

necessity "out there". If the analytical philosophers are not satisfied with this, Ruse tells them that they will have to go away disappointed. The philosophers will presumably not be happy with a viewpoint that threatens to put them out of work, but anyone who has lost patience with the apparently endless wrangling of academic philosophy can only applaud.

On the question of morality, Ruse is critical of "social Darwinism" and other efforts to deduce ethical standards from evolution, precisely because they invariably assume that nature is progressing towards some ultimate good. His own approach starts from the principles of sociobiology, but attempts to purge the remnants of progressionism from the work of Wilson and others. Natural selection will have endowed us with epigenetic rules inclining us to behave in a socially responsible manner, when this is to our (reproductive) advantage. The special status we accord to moral laws is nature's way of ensuring that we have a strong tendency to obey them even when they violate our immediate preferences. As a result, we *feel* that moral laws are objective — and our desire to live in an orderly society ensures that we have no interest in rejecting the laws even when we realize that this feeling is an illusion. Once again, Ruse identifies Hume as the philosophical progenitor of this position.

It will be interesting to see how the opponents of sociobiology respond to this book. Far from justifying a harsh social policy, Ruse explains why we feel so strongly about traditional moral values. Those who wish to argue that human beings are infinitely malleable by culture will object to this — but in so doing they will have to accept that we have no built-in moral values at all. If we concede that our evolutionary heritage may have imposed at least some constraints on how we think and behave, Ruse offers us a way of trying to understand the character of such restraints. As we move into a world that differs ever more radically from that in which our ancestors evolved, such insights may help us to come to terms with the philosophical problems that confront us. □

Peter J. Bowler is a Lecturer in the Department of History and Philosophy of Science, Queen's University of Belfast, Belfast BT7 1NN, UK.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Work of worms — section through a fallen Druidical stone at Stonehenge, indicating how much it has sunk into the ground through the undermining action of earthworms. Depth of the stone is c. 30 cm. The illustration is taken from a new paperback edition of Charles Darwin's *The Formation of Vegetable Mould, through the Action of Worms* (1881), published by University of Chicago Press. Price is \$12, £10.25. For review see *Nature* 24, 553–556 (1981).

New journals review



On 25 September *Nature* will publish the sixth annual review supplement devoted to science journals.

Criteria for inclusion of a journal in the 1986 issue are that:

(i) the first number appeared, or the journal was retitled, between June 1984 and May 1985 (the second cut-off date allows at least three issues of a journal to have been published, the minimum number on which a reasonable judgement can be based);

(ii) it is published at least three times a year;

(iii) the main language used is English.

Publishers and learned societies are invited to send four different issues of each suitable periodical, including the first and most recent numbers (if from outside the United Kingdom, by air mail) to: The Review Editor, *Nature*, 4 Little Essex Street, London WC2R 3LF, England. Subscription details for both 1986 and 1987 should be included.

RÖNNELLS

Scientific Books & Periodicals

For current catalogues:

No. 55

**HISTORY OF
MEDICINE**

No. 56

AMERICANA

and other subjects

write:

RÖNNELLS

Birger Jarlsg. 32
114 29 Stockholm

Parallel promises in computing

Igor Aleksander

The Connection Machine. By W. Daniel Hillis. MIT Press:1986. Pp.190. \$22.50, £22.50.

THE jacket of this book heralds "the connection machine" as "the second state in the evolution of digital computers"; the first sentence reads "Someday, perhaps soon, we will build a machine that will be able to perform the functions of the human mind, a thinking machine". It must be said immediately that while the concept of the connection machine is intriguing, and the book tells us a lot about the engineering behind it, any reference to "thinking machines" is thoroughly misleading.

The primary characteristic of the connection machine is parallelism. The prototype discussed by Hillis contains approximately 64,000 simple processors that may be linked in large parallel groups, and it is one of the strengths of the book that the author discusses his design decisions with great clarity. The main attraction of the machine is that it allows a user to research the behaviour of different "shapes" of parallel machines; for example, the properties of processors connected in a ring are very different from those of processors connected in a matrix, and the connection machine may be configured into these structures and many more. It almost becomes possible to believe the claim that the machine marks a new era in digital computing, as the possibility of working with different "shapes" of parallel structure is awkward (that is, slow) on conventional computers.

Hillis's main problem, in common with

that of designers of similar machines, was to devise schemes that allow processors to be selected from a pool, and then to provide a sufficiently rich set of pre-defined connections so that the user's desired structure can be realized. Clearly, alternatives had to be found to the obvious solution of connecting every processor to every other one: that would require just too many connections, most of which would be unused during the execution of a particular task. Hillis lucidly reviews the various options and describes why he settled on a "Boolean n -cube topology" (the details of which might interest the topologist as an engineer's choice among theoretical alternatives). The end result of the parallel/serial compromise is an achievement of 1,000 million instructions per second (Mips) — the average modern machine in a research laboratory probably delivers about one Mip. Another achievement is that this machine may be programmed by means of an extension of LISP, easily mastered by those who use the language anyway. Hillis goes to some pains to argue that this is an arbitrary choice and advanced features of the machine could equally be applied to other languages.

But what scientific advances are likely to follow from the application of the connection machine? As is often the case with innovators in engineering, Hillis is less comfortable with the answer to this question than he is with the technical description of his machine. He points principally to questions arising in research on artificial intelligence, and suggests that image-processing could benefit by arranging the processors into two-dimensional arrays. Although this is true, and possibly useful, processor arrays have been built for some time; for example, Michael Duff's CLIP machines at University College London have been doing workmanlike tasks in practical image-processing for about 15 years, and those who have used them have probably discovered most of the principles that such methodology is likely to reveal.

Closer to the heart of AI, Hillis suggests that the connection machine might implement knowledge stores called "semantic networks". Here, two processors might have respective labels such as "lemon" and "sour", with a connection leading from the former to the latter and labelled as "tastes". The methodology was first introduced by Patrick Winston in 1970 and it is a matter of debate whether the lack of applications of such schemes in more recent research is solely due to the lack of a sufficiently parallel machine. Away from AI, Hillis proposes that the connection machine might be used to simulate silicon circuits (very large scale integrated systems), but at the rate of one processor (containing thousands of transistors) for each simulated transistor this has an air of overkill about it.

All this talk of application only occupies a few pages in the first chapter and is evidently not central to the author's interest. Hillis perhaps gets closer to the real potential for the connection machine in the last two pages of the book, where he suggests that through such "granular" computers of ever-decreasing grain-size, computational theory will extend its explanatory powers to an increasing number of natural phenomena. This is where the payoff may really be found. The natural phenomena which, in my opinion, are most in need of attention are those concerning brain-like networks of cells, and the connection machine is an ideal simulation tool for such work. But despite the references to "thinking machines", Hillis does not actually discuss this possibility. The bibliography, although extensive, mostly bears little relation to the text and is notable for its omissions. The most serious of these is notice of the modelling of brain-like systems that has been pursued by Geoffrey Hinton at Carnegie Mellon University since the late 1970s. Exciting discoveries are being made about the way in which cellular learning systems can learn to "understand" concepts such as the shape of objects "seen" through a television camera, or build internal representations of "Italian name" against "English name" in a telephone directory database *without* having been programmed to do so directly. Investigations of this kind have much to do with what is meant by the word "thinking" and, if carried out through the connection machine, will contribute a great deal to understanding the functions of the brain. Indeed, simulations of Hinton's work are clumsy and counter-productive on conventional machines, and it is in this sense that the connection machine provides a useful tool with which "thinking" may be researched. To confuse this with the creation of a "thinking machine" is a category mistake of textbook quality. Moreover, despite the author's implication that the MIT-associated connection machine is unique, it must be seen as taking its place among a number of similar developments such as the "Non-Von" machine at Columbia University and John Darlington's ALICE machine at Imperial College, London.

The Connection Machine will be appreciated best by those who require a lucid description of a computer design that is truly ingenious and indicative of the trend in computer science hardware research towards massive parallelism. It is not for those with no knowledge of computer engineering at all, nor for others who hope to be enlightened on what a "thinking machine" might be. □

Igor Aleksander is Kobler Professor of Information Technology, Department of Computing, Imperial College, University of London, London SW7 2BZ, UK.

The real Frederick Soddy*



*As a number of people have pointed out, the photograph supporting the review of the collection of essays on Soddy (*Nature* 320, 691; 1986) was in fact of Ernest Rutherford.

Muon-catalysed fusion revisited

Steven Earl Jones

Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA

Muons introduced into relatively cold, dense deuterium-tritium mixtures can replace the atomic electrons and form muonic molecules which participate readily in nuclear fusion reactions. Catalysis yields of ~150 fusions per muon have been achieved, renewing interest in muon-catalysed fusion as a possible source of energy.

MUON-CATALYSED fusion can be viewed as a series of reactions induced by negative muons in cold hydrogen leading to nuclear fusion and a release of energy. For example, when a negative muon enters a mixture of deuterium and tritium (heavy isotopes of hydrogen):



The muon replaces an electron to form a $t\mu$ atom:



A molecule bound by the muon may then form:



which leads to deuterium-tritium (d-t) fusion, freeing the muon to repeat the cycle:



Occasionally the muon is captured by the α -particle following fusion:



The key to this remarkable sequence of reaction lies in the large mass of the muon, which is 207 times that of the electron. A negative muon admitted into a high-density mixture of deuterium and tritium gases (step 1) quickly replaces the lighter electron to form a muonic atom (step 2). The reaction is energetically favourable and rapid because the muon greatly outweighs the electron. The same holds true for the succeeding step (3), formation of a $dt\mu$ molecule held together by the muon. The muonic molecule is ~200 times smaller than an ordinary heavy hydrogen molecule (such as DT, bound by electrons). Thus, the muon confines the d-t nuclei to a very tiny volume. Although the rest of the gas may be near room temperature, conditions inside the muonic molecule are similar to those found inside a white dwarf star¹. In such conditions nuclear fusion (step 4) will occur, just as it does in stars. Indeed, d-t fusion occurs in $\sim 10^{-12}$ s in the $dt\mu$ molecule².

Without the use of external magnetic fields or lasers, we can thus achieve local star-like conditions by simply adding negative muons to a mixture of hydrogen gases. Each muon can spontaneously induce a number of fusion reactions, so that it serves as a catalyst for nuclear fusion.

In his 1968 Nobel prize acceptance lecture, Luis Alvarez³ described the excitement generated by the first observation⁴ in 1956 of this fascinating phenomenon for muons stopped in liquid hydrogen (H) plus deuterium (D):

"We had a short but exhilarating experience when we thought we had solved all of the fuel problems of mankind for the rest of time. A few hasty calculations indicated that in liquid HD a single negative muon would catalyse enough fusion reactions before it decayed to supply the energy to operate an accelerator to produce more muons, with energy left over after making the liquid HD from sea water. While everyone else had been trying to solve this problem by heating hydrogen plasmas to millions of degrees, we had apparently stumbled on the solution, involving very low temperatures instead."

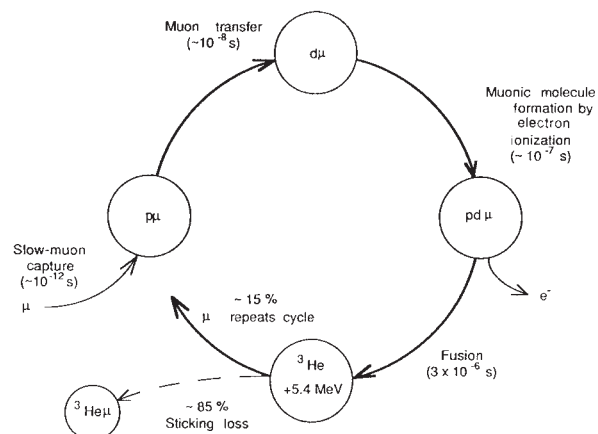


Fig. 1 The muon catalysis cycle for liquid hydrogen containing 1% deuterium. Approximate reaction times are shown^{8,9}. Because the mean time required for muon-induced fusion to occur in the $pd\mu$ molecule (3 μ s) exceeds the mean muon lifetime (2.2 μ s), the muon usually decays before it can repeat the $pd\mu$ fusion cycle.

Excitement quickly waned, however, following theoretical predictions that very few fusions could be catalysed by a single muon. Indeed, in experiments performed by Alvarez and others using hydrogen and deuterium, only rarely was a muon observed to induce more than one fusion reaction.

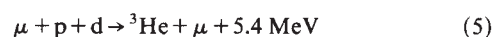
In mixtures of deuterium and tritium, the situation is strikingly different, for reasons which are as yet only partially understood. Within the past few years, we have demonstrated that a single muon can complete the reaction sequence (1)–(4) many times, catalysing many d-t fusions. It is fitting on this thirtieth anniversary

announce the achievement at the Los Alamos Meson Physics Facility of 150 fusions per muon (average) in dense deuterium-tritium mixtures⁵. These yields exceed theoretical expectations, and have stimulated speculation regarding muon-catalysed fusion as a source of energy.

Muon-catalysed fusion research encompasses many disciplines, including atomic, molecular, nuclear and particle physics, chemistry, accelerator technology, and possibly fusion energetics. This review, which is intended to be accessible to scientists in diverse fields, highlights the most recent discoveries which have stimulated renewed interest in muon-catalysed fusion.

Muon-catalysed p-d fusion

The first type of muon-catalysed fusion to be discussed theoretically^{6,7} and observed experimentally^{3,4} was proton-deuteron (p-d) fusion:



The first observation of muon-induced fusion came as a great

surprise^{3,4}. Alvarez and his colleagues at Berkeley noticed in their liquid-hydrogen bubble chamber a number of mysterious muon tracks having a characteristic 1.7-cm length. Furthermore, a gap frequently appeared between the 1.7-cm track and the stopping point of the incoming muon.

Unaware of earlier theoretical work by Frank⁶, A. D. Sakharov and P. N. Lebedev^{5,4} and Zel'dovich⁷, the Berkeley physicists inferred the phenomenon of muon-catalysed fusion from their data. The principal clue was the 1.7-cm range of the muons, which corresponds to an energy of 5.4 MeV—just the energy released in proton-deuteron fusion. It was quickly realized that a muon could form a muonic ($d\mu p$) molecule, and then, “in such a ‘molecule’, the proton and deuteron would be brought together in such a close proximity for such a long time that they would fuse into ${}^3\text{He}$, and could deliver their fusion energy to the muon by the process of internal conversions”³.

The mystery of the observed gap between the stopped and ejected muons was solved by Edward Teller^{3,4}. He correctly surmised that the muon would be initially captured by a proton, the nucleus of the predominant isotope of hydrogen in the bubble chamber, to make a muonic $p\mu$ atom. Although the fraction of deuterons in naturally occurring hydrogen is only 1 in 5,000, the muon could transfer rapidly and irreversibly to a deuteron to make a $d\mu$ atom:

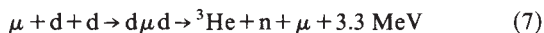


because the binding energy for the ground-state $d\mu$ atom is 135 eV greater than that of the $p\mu$ atom. The neutral $d\mu$ atom could then recoil some distance, producing the observed gap, followed by $d\mu p$ molecule formation and fusion. The ability of the $d\mu$ atom to travel a few millimetres in liquid hydrogen before stopping (that is, before becoming thermalized) provides an interesting example of the Ramsauer-Townsend effect⁸.

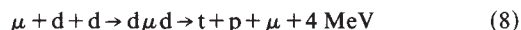
The sequence of reactions just discussed is shown in Fig. 1, along with some observed reaction times^{8,9}. A major limiting factor in the process is the likelihood that the negative muon will be captured by the positive ${}^3\text{He}$ ion synthesized during proton-deuteron fusion; another is the long time spent by the muon in the $p\mu$ molecule before fusion occurs. These factors limit the number of muon catalysis cycles in liquid hydrogen containing $\sim 1\%$ detuterium to an average of less than one fusion per muon.

Muon catalysis in pure deuterium

The Alvarez group and others⁸ extended the study of muon catalysis to the case of pure liquid deuterium, observing the reactions



and



Experimental results, shown in Fig. 2, were consistent with expectations based on studies of hydrogen-deuterium mixtures. The relatively slow formation rate of the muonic molecule and the loss of muons due to ${}^3\text{He}$ ion capture limited the average number of fusions per muon to less than one. It thus appeared that muon-catalysed fusion reactions in hydrogen isotopes were rather well understood and quite useless as an energy source^{8,10,11}.

Persistent research in the Soviet Union, however, led to a revival of interest in muon catalysis¹². Dzhelepov *et al.*¹³ first showed experimentally that the rate at which $dd\mu$ muonic molecules form depends strongly on temperature. In fact, the $dd\mu$ formation rate in deuterium at room temperature was found to be 10 times greater than the rate at liquid-deuterium temperatures (~ 30 K). That the deuterium temperature could strongly affect the formation rate of $d\mu d$ molecules came as a complete surprise.

In 1967, following an earlier suggestion of Zel'dovich⁷,

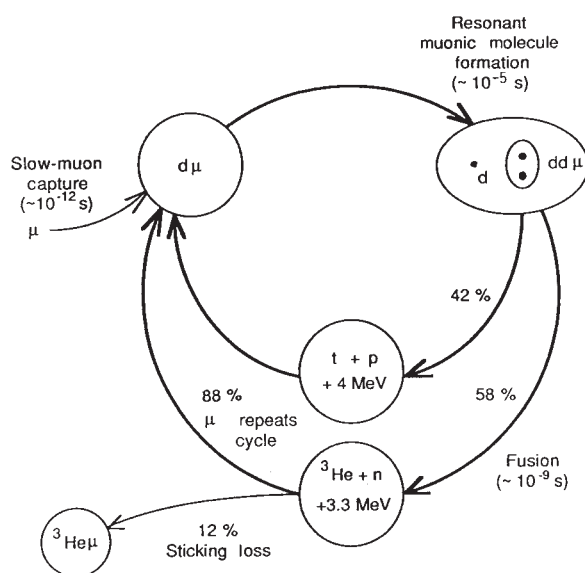


Fig. 2 The major reactions which occur when a muon is admitted into liquid deuterium. Resonant formation of a $dd\mu$ molecular ion within an electronic molecule (without emission of an electron) can occur when a $d\mu$ atom penetrates a D_2 molecule. However, the $dd\mu$ molecular formation process is slow relative to the muon lifetime so that the muon often does not survive to repeat the $dd\mu$ catalysis cycle.

Vesman¹⁴ advanced the notion that the muonic $d\mu d$ molecule could form within a D_2 host molecule, as shown in Fig. 2. Note that the $(d\mu d)^+$ ion serves as one of two nuclear centres in the complex new molecule. For this process to occur, the energy released as the muonic molecule forms ($d\mu d$ binding energy, E_b , plus the initial kinetic energy of the $d\mu$ atom, E_k) must be small enough to be absorbed in the vibrational and rotational excitation ΔE of the host molecule:

$$E_b + E_k = \Delta E \quad (9)$$

Because the energy levels of the complex molecule are quantized, equation (9) expresses a resonance condition; that is, the muonic molecule forms most rapidly when equation (9) is exactly satisfied. The temperature dependence of the process derives from the E_k term.

Precise calculations showed that there exists an excited state of the muonic $d\mu d$ molecule that has a binding energy of only 1.9 eV, which is small enough to be absorbed by the electronic molecule without causing it to break apart. Furthermore, the temperature dependence of the $d\mu d$ formation rate could be accounted for by the resonance requirement^{15,16}.

A startling prediction followed from this theoretical model: the formation of muonic $dt\mu$ molecules should also be resonant, and resonant $dt\mu$ formation at 540 K (the expected optimum temperature) should exceed the fastest $d\mu d$ formation rate by two orders of magnitude¹⁷. In addition, muon capture by the helium ion synthesized during deuteron-triton fusion had been predicted to be $\sim 0.9\%$ ¹⁰, much less than the corresponding loss term for proton-deuteron or deuteron-deuteron fusion. Thus, in late 1977 Gershtein and Ponomarev¹⁷ declared:

“It means that every μ^- meson in the mixture deuterium and tritium can produce $\sim 10^2$ of the nuclear fusion reaction and release ~ 2 GeV which is 20 times the rest mass of muon. We leave the question open about the possibilities of practical applications of the phenomenon discovered . . .

The reactions [in deuterium plus tritium] are not yet studied experimentally. We hope that such experiments will soon be performed despite the technical difficulties involved.”

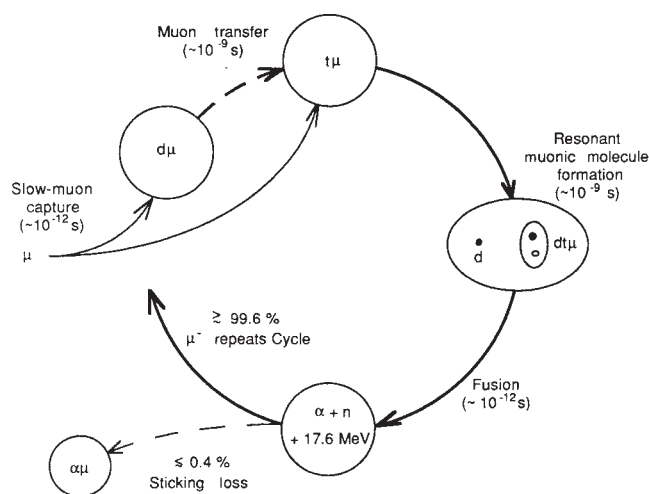


Fig. 3 The major processes which occur when a muon stops in a liquid mixture of deuterium and tritium in roughly equal proportions. The reaction times are very short, compared with the muon lifetime (2.2 μ s). In addition, the probability of muon retention by the α -particle synthesized during d-t fusion is small. Thus, the muon-catalysed d-t fusion cycle can easily be repeated more than 100 times.

Experiments were soon mounted. As we shall see, some results were as startling as the predictions. Muon-catalysed fusion was not a closed subject after all.

Muon catalysis in D-T mixtures

Resonant $dt\mu$ formation. Bystritsky *et al.*^{18,19} set out in 1979 to test predictions that the $dt\mu$ molecule could form in $\sim 0.5\%$ of the muon lifetime and that this formation rate would increase with increasing deuterium-tritium (D-T) gas temperature. They showed that $dt\mu$ molecules may form at least as fast as predicted by the new resonance model (shown in Figs 2 and 3), and four orders of magnitude faster than expected from the electron-emission process (shown in Fig. 1). The experiment allowed only a lower limit to be set for the $dt\mu$ formation rate.

However, no temperature dependence of the $dt\mu$ formation process was detected, even though the D-T gas was cooled to 93 K and heated to 613 K. Rapid $dt\mu$ formation appeared consistent with the resonance model, but where was the expected temperature dependence? Unfortunately, the Dubna accelerator was closed for refurbishment in 1979 and the Soviet experimenters were unable to pursue their research. (The machine is expected to return to operation soon. Related experiments recently began at the Gatchina facility of the Leningrad Nuclear Physics Institute.)

An experiment was begun in 1982 at the Los Alamos Meson Physics Facility (LAMPF) by our group, with the intention of exploring in depth the muon catalysis cycle in D-T mixtures^{5,20-23}. Vessels capable of holding deuterium and tritium gases at pressures up to 1,000 atm (later, 3,000 atm) were developed at the Idaho National Engineering Laboratory (INEL) to allow near-liquid-hydrogen densities to be maintained over a broad temperature range (13–800 K). The experimental team consists of scientists from Brigham Young University (BYU), INEL and the Los Alamos National Laboratory (LANL).

Early one morning in November 1982, we cooled a target containing equal amounts of deuterium and tritium from room temperature to 100 K: the number of fusion neutrons decreased significantly as the gas was cooled. Although in qualitative agreement with the resonance model prediction, the result was surprising because no temperature dependence had been detected

in the earlier Dubna experiment. The rest of the night was spent checking the electronics for a malfunction, but nothing was wrong. As the D-T mixture was heated the following day, the fusion yield grew larger. Resonance model predictions of temperature dependence enunciated by Ponomarev and colleagues were finally confirmed.

However, virtually no temperature dependence was seen when a target composed of 10% tritium and 90% deuterium was used. This result was consistent with the Dubna experiment, in which the D-T mixture contained $\leq 8\%$ tritium. The experiments soon demonstrated that for low tritium fractions, the muon catalysis cycle was dominated by the muon transfer reaction



This reaction, analogous to reaction (6), is nearly independent of temperature. These two facts explain why no temperature dependence was seen in the low-tritium-fraction experiments at Dubna. For tritium fractions $\geq 30\%$ at LAMPF, the temperature-dependent $dt\mu$ formation rate became evident^{5,20-23}.

With several hundred million muon-catalysed fusion events now recorded in the continuing BYU/INEL/LANL experiments, along with continued theoretical efforts, a consistent picture of the resonant formation of $dt\mu$ molecules is emerging. The important features are:

(1) The $dt\mu$ -formation rate rises with increasing temperature, as expected²⁰⁻²³.

(2) The rate does not peak at ~ 540 K as predicted^{15,16}. Instead, it continues to climb at 800 K, the highest temperature reached to date. There are now theoretical reasons to expect the rate to reach a maximum near 1,200 K²⁴. Note that the resonance process depends on the presence of intact molecules to host the $(dt\mu)^+$ ion (see Fig. 3). As hydrogen molecules begin to decompose into atoms at temperatures $\geq 5,000$ K, the resonance formation mechanism becomes unavailable so that $dt\mu$ formation must decline precipitously. Moreover, muon-catalysed fusion cannot be used in conjunction with plasmas; in particular, it cannot be used for thermonuclear weapons. The term 'cold fusion' is therefore quite appropriate for the process.

(3) The $dt\mu$ formation rate depends on whether the host is a D_2 or DT molecule^{5,20-23}. In particular, the reaction

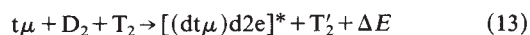


proceeds more rapidly over a broad range of density and temperature than does the reaction



This host-dependence arises from the different energy levels which are available in the two compound molecules, in keeping with the resonance model²⁴.

(4) While the rate of reaction (12) scales linearly with density as anticipated, the rate of reaction (11) increases faster than linearly with density⁵. The resonance model implies that $t\mu + D_2$ collisions are special, in having their strongest resonances just below threshold. A third body can allow these resonances to contribute by absorbing some kinetic energy in a reaction such as



Thus, the density dependence of reaction (11) observed at LAMPF provides evidence for significant resonant $dt\mu$ formation involving three-body collisions^{5,25}, although other explanations for the phenomenon cannot be ruled out²⁶.

Muon capture by the fusion α -particle. Experiments thus demonstrate that $dt\mu$ molecules do form extremely rapidly, nearly three orders of magnitude faster than $pd\mu$ or $dd\mu$ molecules. In fact, owing to the extraordinary resonance in $dt\mu$ formation, a muon could catalyse many hundreds of d-t fusions during its lifetime, were it not for muon capture by the synthesized helium nucleus (see equation (4b)). We have seen that muon loss in this way constitutes a dead-end in the p-d and d-d catalysis

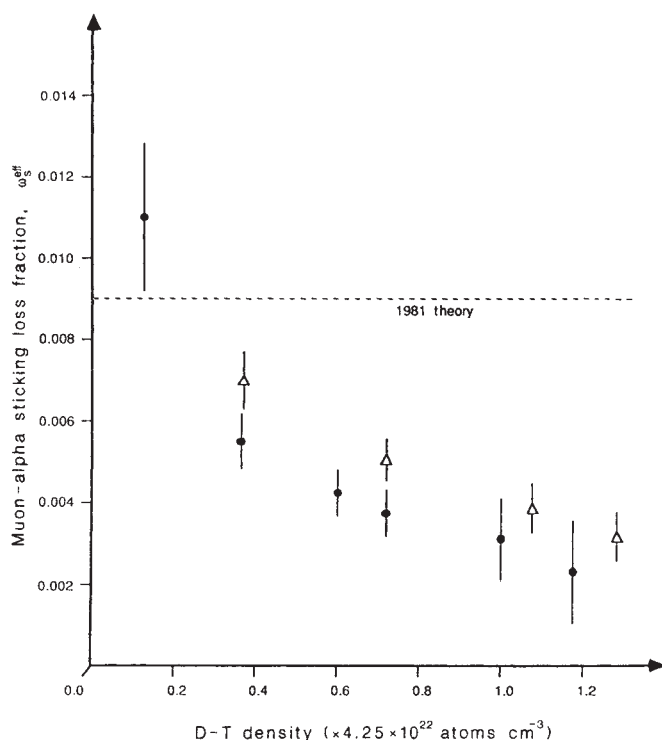


Fig. 4 Data from LAMPF showing the decrease in the deduced μ - α sticking fraction as the density of the deuterium-tritium mixture is increased, for 50% tritium/50% deuterium (dots) and 30% tritium/70% deuterium (triangles). At high target densities, the μ - α sticking probability is clearly much lower than was predicted before the experiments were performed.

cycles (Figs 1 and 2); does this feature persist in d-t fusion?

Before 1985, theorists^{1,10,27} predicted that the muon would be captured by the fusion-synthesized α -particle in $\omega_s^0 = 1.2\%$ of muon-induced fusion reactions. Then, as the $\alpha\mu$ ion moves rapidly through the surrounding gas, there is a probability $R \approx 24\%$ that the muon will be collisionally stripped from the α particle and thereby freed to continue the cycle. Thus, after slowing down in the gas, the muon would remain bound to the α -particle with probability $\omega_s = \omega_s^0(1 - R) \approx 0.9\%$. Muon loss due to capture by the synthesized α -particle is cumulative, so that the maximum number of d-t fusions which a muon could possibly catalyse would be $1/\omega_s \approx 110$ (ref. 10). This limit would apply even if the muon were stable. Thus, the μ - α sticking probability ω_s was perceived to be the ultimate limit on the fusion yield from the muon catalysis cycle^{1,10,17}.

The only published measurements of ω_s are from the BYU/INEL/LANL experiments^{5,20-23}. The deduced quantity ω_s^{eff} (expected to represent ω_s (ref. 5)) is shown in Fig. 4; the results were totally unexpected. Note the striking dependence of the apparent μ - α sticking loss fraction on the density of the D-T mixture. Moreover, the magnitude of ω_s at high gas density is approximately one-third the theoretical prediction of 0.9%. A possible dependence on tritium fraction is also suggested by the data but is statistically inconclusive.

Neither the smallness of ω_s nor its apparent density dependence is understood. Perhaps the muon stripping fraction R depends strongly on density and attains large values for dense D-T mixtures. Then $\omega_s = \omega_s^0(1 - R)$ could become small for high densities. Although theorists are now re-examining muon stripping probabilities, this explanation does not appear promising. Or perhaps ω_s^0 itself varies with density; for example, ω_s^0 may depend on the state of the $\text{dt}\mu$ molecule from which fusion occurs (C.-Y. Hu, in preparation), which in turn could depend

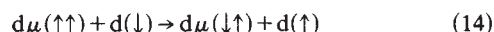
on density. It has also been suggested that a certain fraction of muons may be significantly delayed as the $\text{dt}\mu$ molecule cascades down from the excited state in which it is formed, thus increasing ω_s^{eff} above a small initial value²⁸. Two years after the initial observation of this phenomenon at LAMPF, the puzzle persists. It is appropriate to emphasize that no independent confirmation of the LAMPF results has yet been published.

Interestingly, theoretical estimates of values ω_s have decreased significantly since small ω_s values were reported in 1984^{21,23}. Jackson, who first calculated ω_s^0 using the Born-Oppenheimer approximation in 1957¹⁰, estimated that corrections to that approximation would reduce ω_s^0 by $\sim 25\%$ ²⁹. In agreement with this estimate, a precision Monte Carlo calculation by Ceperley and Alder³⁰ yielded $\omega_s = 0.6\%$. These recent calculations are, however, still larger than the (high-density) experimental values. Mueller and Rafelski³¹ argue that the (resonant) nuclear d-t interaction, neglected in other calculations, is capable of affecting the wave function describing the three-body $\text{dt}\mu$ system. This may have the effect of reducing the initial sticking to $\omega_s^0 \approx 0.1\%$ —a remarkably small value.

Although not yet understood, the BYU/INEL/LANL results clearly point to small ω_s values (Fig. 4). Thus, >300 fusions per muon may be possible in deuterium-tritium mixtures. In the past few years there has been a steady rise in the average number of d-t fusions per muon achieved at LAMPF: 20 in 1982, 80 in 1983 and 150 in 1984. Each step reflects an increased understanding of the rich physics underlying muon-catalysed fusion.

Hyperfine effects. In addition to the research in the Soviet Union and the United States, an experiment on muon-catalysed d-t fusion is in progress at the Swiss Institute for Nuclear Research (SIN). The work by Breunlich *et al.*³² has so far focused on the fascinating possibility of hyperfine-structure effects.

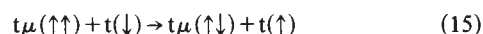
It was recognized long ago that the reactions in muon catalysis could be influenced by the spin states of muonic atoms and molecules. Indeed, experiments at Columbia University³³ in the early 1960s showed that hyperfine transitions take place in muonic atoms, according to the reaction



By increasing the deuterium concentration in a liquid hydrogen target from 0.7 to 25%, Lederman and collaborators enhanced the population of $\text{d}\mu(\downarrow\uparrow)$ atoms. This in turn increased the population of those $\text{pd}\mu$ molecules in which fusion is most rapid, and the number of proton-deuteron fusions per muon increased by $\sim 17\%$ (Wolfenstein-Gershtein effect)³³. Of course, the average yield was still less than one p-d fusion per muon.

Interest in hyperfine effects, as in muon catalysis generally, was stimulated by the discovery of the resonant formation of $\text{dd}\mu$ and $\text{dt}\mu$ molecules discussed above. Breunlich and co-workers repeated studies of hyperfine effects in hydrogen-deuterium mixtures⁹, and then observed hyperfine effects in pure deuterium³⁴. In particular, they found that the rate of resonant $\text{dd}\mu$ formation in $\text{d}\mu + \text{D}_2$ collisions depends strongly on the hyperfine state of the $\text{d}\mu$ atom. The experimental results can be described within the framework of the resonance formation model for $\text{dd}\mu$ molecules. Indeed, a rather complete picture of the energy levels of the muonic $\text{dd}\mu$ molecule, including hyperfine splitting, is anticipated from these experiments.

However, results of experiments conducted at SIN with mixtures of deuterium and tritium appeared to contradict theory. Time spectra of the fusion neutrons gave evidence of two distinct $\text{dt}\mu$ formation rates³². (This effect is also reported in the LAMPF experiments²².) When the SIN experimenters interpreted their data in terms of hyperfine effects, the results were contrary to theoretical expectations. For instance, the evaluated rate of the spin-flip reaction



was much smaller than predicted and displayed an unexpected and strong dependence on temperature. Moreover, very little

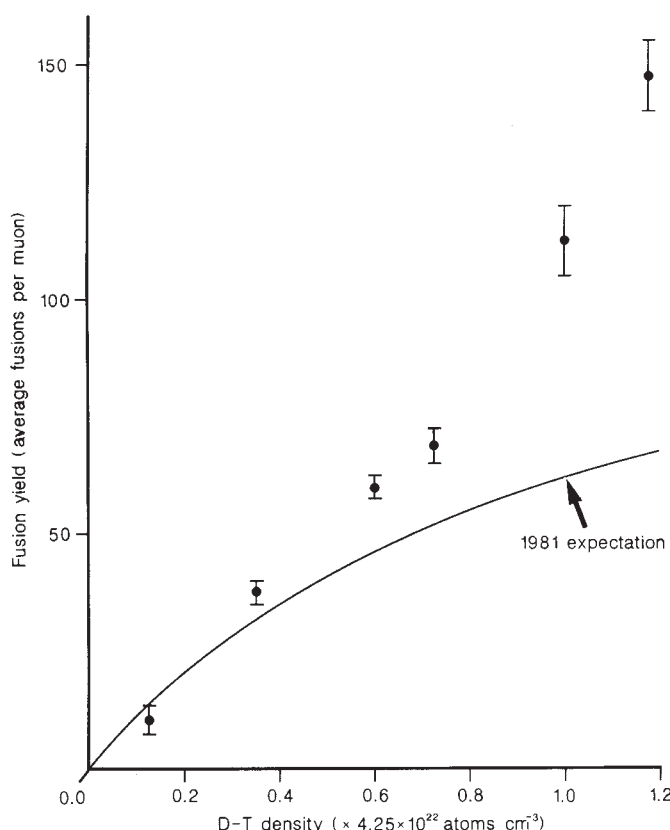


Fig. 5 The average number of muon-catalysed d-t fusion cycles observed by the BYU/INEL/LANL collaboration as a function of target density for cold ($T \leq 100$ K), equimolar deuterium-tritium mixtures. The solid line is based on experimental data^{18,19} and theoretical predictions³⁷ as of 1981.

temperature dependence of the $dt\mu$ formation rate emerged from the analysis of the SIN data³².

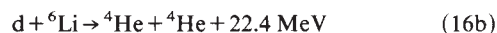
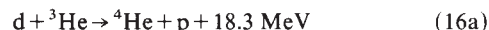
An alternative explanation has been recently advanced^{24,35}; perhaps the data reflect different $dt\mu$ formation rates for thermalized or 'stopped' $t\mu$ atoms than for 'hot' recoiling $t\mu$ atoms. This is reminiscent of the way in which the recoil of $d\mu$ atoms accounted for the peculiar gaps observed in the Alvarez experiments. In the present case, a strong resonance for 'hot' $t\mu$ atoms colliding with D_2 or DT molecules would account for a very large $dt\mu$ formation rate. (This suggests the interesting possibility that if the gas were heated to $\sim 1,200$ K, $dt\mu$ formation would occur extremely rapidly²⁴.) The 'hot atom' model accounts for the observed density dependence of the neutron time spectra²², among other features, and is now generally favoured over the hyperfine-effects interpretation. With this explanation, the fusion-neutron time spectra observed at LAMPF and SIN appear consistent with the resonant $dt\mu$ molecule formation model. Clearly, hyperfine effects will be more difficult to extract in $dt\mu$ catalysis than in $dd\mu$ and $pd\mu$ catalysis. In any case, only the lower-energy hyperfine state $t\mu(\uparrow\downarrow)$ should be significantly populated for tritium fractions $\geq 10\%$ and densities $\geq 0.1 \times$ liquid-hydrogen density, because of the rapidity of the spin-flip reaction (15)³⁶.

Summary. The physics of muon catalysis in mixtures of deuterium and tritium is unexpectedly rich. The major processes and experimental values now thought to describe muon-catalysed d-t fusion are displayed in Fig. 3. By comparing Figs 1-3, one sees why the fusion yield for the muon-catalysed d-t fusion cycle greatly exceeds fusion yields in the p-d and d-d catalysis cycles. The average number of d-t fusion cycles per muon measured at LAMPF by the BYU/INEL/LANL collaboration is shown in Fig. 5. The observed yields at high D-T mixture

densities exceed even the optimistic (resonance model) predictions³⁷, reflecting the discovery of a very small μ - α sticking probability⁵.

Other reactions

Can muons induce fusion reactions involving elements heavier than hydrogen? The following reactions, for example, are appealing because there are no neutrons in the final state:



A muonic molecule must form before muon catalysis can proceed^{8,10} and the rates for $d\mu$ and $d\mu$ formation have been calculated to be rapid relative to muon decay³⁸⁻⁴⁰. However, Coulomb repulsion in such cases is large, as is the nuclear separation, so the fusion rate is small^{10,37}. The normal fate of the muon is to become irretrievably bound to the heavier isotope as the molecules dissociate in picoseconds³⁸⁻⁴⁰.

Muon catalysis experiments involving hydrogen isotopes contaminated with ${}^3\text{He}$ (refs 20-23) or ${}^4\text{He}$ (refs 41, 42) provide a test of these predictions. Helium isotopes, are, in fact, observed to scavenge muons rapidly, thus poisoning the muon catalysis cycle. Consequently, it is desirable to keep helium and other non-hydrogen impurities at a low level⁸. It is intriguing that muon-catalysed fusion proceeds rapidly only for deuterium-tritium mixtures.

Energy applications

In experiments performed by the BYU/INEL/LANL collaboration at LAMPF, an average fusion energy release of ~ 3 GeV per muon has been achieved and larger yields are expected⁵. The achieved fusion energy yield is 25 times greater than the total energy of the muon which drives the fusion reactions. However, muon-catalysed fusion as an energy source suffers from the lack of an efficient way to generate the muon catalysts. It appears that ~ 5 GeV must be invested to produce one negative muon by advanced techniques⁴³⁻⁴⁸.

In 1980, Petrov^{43,44} presented a clever scheme for a power reactor based on muon-catalysed fusion combined with nuclear fission processes and found that a commercial reactor would require ~ 100 fusions per muon. Others have examined and amplified the fusion/fission hybrid scheme and concluded that 100-300 fusions per muon would suffice for a power plant⁴⁵⁻⁵⁰. The LAMPF experiments demonstrate that such yields are indeed possible (see Fig. 5). A pure fusion reactor might require 1,000 fusions per muon, based on current ideas for efficient muon generation⁵¹.

Figure 6 shows a schematic outline of a conceptual muon-catalysed fusion system based on the recent work of Eliezer *et al.*⁵⁰. The first step is to form a beam of high-energy tritium nuclei using an efficient particle accelerator; the development of the radio-frequency (r.f.) quadrupole during the last decade represents a relative advance in accelerator technology. Grid-to-beam power conversion efficiencies as high as 60% appear achievable, along with high beam intensities (~ 200 mA)⁴³. Dramatic progress in superconducting r.f. cavities for particle accelerators⁵² may also find application here.

The high-energy (~ 1 GeV/nucleon) triton beam enters one or more pressurized D-T gas targets. Negative pions are copiously produced in ensuing nuclear collisions and these rapidly decay into muons. (The ratio of negative-to-positive pions and muons is enhanced through the use of neutron-rich tritium nuclei in the beam and the target^{43,44,49}.) Eliezer *et al.*⁵⁰ calculate pion/muon production and stopping rates in the D-T gas; a magnetic mirror effectively confines the charged particles to a small volume (~ 0.1 m³). The muons are produced and stopped in the D-T fuel itself, which proves to be much simpler and more efficient than forming intermediate pion and muon beams as is conventionally done at accelerator laboratories, even

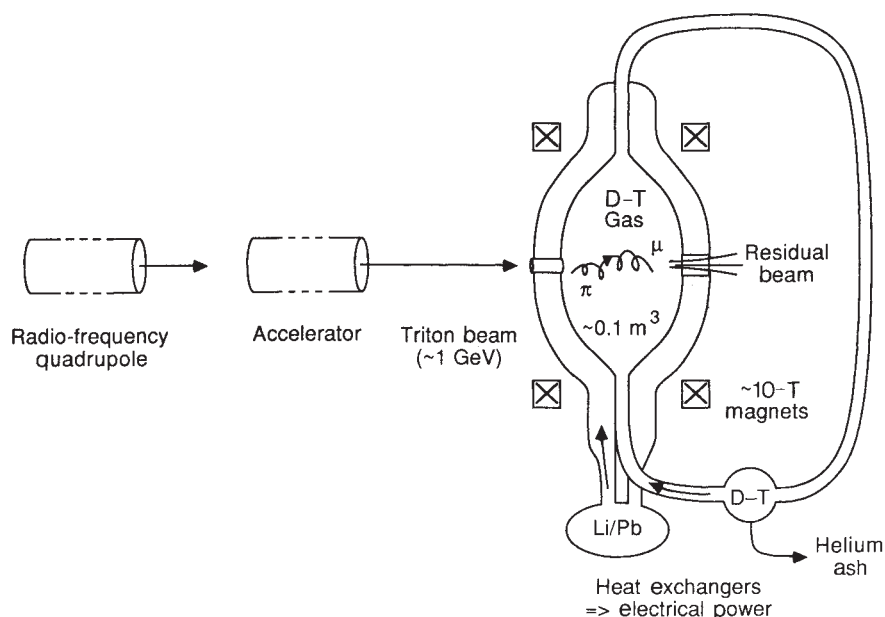


Fig. 6 A conceptual muon-catalysed fusion system, showing detail of a modular reaction vessel where muon-catalysed d-t fusion occurs.

though about one-third of the pions are lost by nuclear capture processes in the D-T gas.

Muons then spontaneously and rapidly catalyse fusion reactions in the D-T fuel as we have discussed, producing energetic α -particles and neutrons. The α particles stop within ~ 1 mm in the dense gas which protects the vessel walls from α -bombardment. The gas must be circulated to extract heat and ^4He , the inert product of the fusion reaction. Note that in a reactor environment, thermal neutrons would consume ^3He arising from tritium decay by the reaction ($^3\text{He} + n \rightarrow p + t$)⁵³.

The energetic neutrons enter a lithium-lead eutectic melt, where they deposit heat and produce tritium by the reactions ($n + ^6\text{Li} \rightarrow \alpha + t$) and ($n + ^7\text{Li} \rightarrow n + \alpha + t$). A fertile isotope such as thorium could be used in the blanket to greatly amplify the energy gain from the muon-catalysed fusion neutrons⁴³⁻⁵⁰.

A few researchers have studied ways of recovering the significant energy remaining in the form of energetic protons and neutrons following muon production⁴³⁻⁵⁰. One approach is to use the residual beam to breed fissile fuel; for example, spent fuel rods from fission reactors could be renewed without chemical reprocessing. Alternatively, the energy of charged particles could be recovered directly; neutrons would first produce energetic protons through inelastic collisions (G. Chapline, personal communication). Alvarez (personal communication; see also ref. 49) and Eliezer *et al.*⁵⁰ have independently suggested that beam particles which do not interact strongly with the target could be collected, re-accelerated and re-circulated into the reaction chamber.

The scheme of Fig. 6 can be characterized as a power amplifier, with input power to the accelerator ultimately driving the muon-catalysed fusion reactor. There is neither plasma to heat nor danger of thermal runaway. However, it is clear that for such a scheme to succeed: (1) hundreds of fusions per muon must be achieved; (2) muons must be produced cheaply (<10 GeV per muon); and (3) many engineering problems must be solved. Research now centres on the first issue: we are surprised to observe so many catalysed d-t fusions per muon (see Fig. 5) and we are exploring the limits of the process. A better understanding of the physics may suggest ways to further enhance the cold fusion process. The research could thus lead to the serious consideration of muon-catalysed fusion as an energy source.

Conclusion

A resonance which allows muonic $\text{dt}\mu$ molecules to form in less than one-thousandth of the muonic lifetime is indeed a remarkable phenomenon. The resonant process has been

observed when negative muons are stopped in 'cold' mixtures of deuterium and tritium gases and leads to rapidly induced nuclear fusion reactions. This effect could hardly have been predicted from the studies of muon reactions in liquid hydrogen and deuterium 30 years ago. The discovery of this resonance in the last few years has reawakened interest in muon-catalysed fusion.

On top of this surprise comes another: the retention of muons by α -particles synthesized during d-t fusion is much less severe than expected. Muon catalysis research now focuses on confirming and understanding this discovery. Together these phenomena imply that each muon may catalyse hundreds of d-t fusion reactions, releasing a great deal of fusion energy. Not surprisingly, these discoveries, along with advances in accelerator technology, have renewed interest in muon-catalysed fusion as a potential energy source. One might say that muon-catalysed fusion is an idea whose time has come—again.

My colleagues deserve recognition for years of dedicated effort: Alan Anderson, James Bradbury, Gus Caffrey, James Cohen, Peter Gram, Mel Leon, Richard Maltrud, Michael Paciotti, Clinton Van Siclen and Kenneth Watts. I particularly acknowledge also the support of early pioneers in muon catalysis research, Luis Alvarez, David Jackson, David Judd, and Leon Lederman, and of recent bold thinkers such as Leonid Ponomarev, Johann Rafelski and Marshall Rosenbluth. Finally, I thank Ryszard Gajewski of the US Department of Energy, Division of Advanced Energy Projects, for financial support and encouragement.

Note added in proof: Results of the SIN experiment⁵⁵ confirm that μ - α sticking is small: $\omega_s^\circ = (0.45 \pm 0.05)\%$ for liquid D-T mixture, in good agreement with LAMPF result (Fig. 4).

1. Bracci, L. & Fiorentini, G. *Phys. Rep.* **86**, 169 (1982).
2. Bogdanova, L. *et al. Soviet Phys. JETP* **34**, 662 (1981).
3. Alvarez, L. W. *Adventures exp. Phys.* **10**, 72-79 (1972).
4. Alvarez, L. W. *et al. Phys. Rev.* **105**, 1127 (1957).
5. Jones, S. E. *et al. Phys. Rev. Lett.* **56**, 588 (1986).
6. Frank, F. C. *Nature* **160**, 525 (1947).
7. Zel'dovich, Ya. B. *Dokl. Akad. Nauk SSSR* **95**, 493 (1954).
8. Gershtein, S. S. & Ponomarev, L. I. in *Muon Physics* Vol. 3 (eds Hughes, V. W. & Wu, C. S.) 141 (Academic, New York, 1975).
9. Bertl, W. H. *et al. Atomkernenergie-Kerntechnik* **43**, 184 (1983).
10. Jackson, J. D. *Phys. Rev.* **106**, 330 (1957).
11. Zel'dovich, Ya. B. & Gershtein, S. *Soviet Phys. Usp.* **3**, 593 (1961).
12. Rafelski, J. in *Exotic Atoms '79* (eds Crowe, K. & Duclos, E.) 177 (Plenum, New York, 1980).
13. Dzhelepov, V. P. *et al. Soviet Phys. JETP* **23**, 820 (1966).
14. Vesman, E. A. *Zh. eksp. teor. Fiz. Pis'ma Redapt.* **5**, 113 (1967).
15. Vinitsky, S. I. *et al. Soviet Phys. JETP* **47**, 444 (1979).
16. Vinitsky, S. I. *et al. Soviet Phys. JETP* **52**, 353 (1980).
17. Gershtein, S. S. & Ponomarev, L. I. *Phys. Lett.* **72B**, 80 (1977).
18. Bystritsky, V. M. *et al. Phys. Lett.* **94B**, 476 (1980).
19. Bystritsky, V. M. *et al. Soviet Phys. JETP* **53**, 877 (1981).
20. Jones, S. E. *et al. Phys. Rev. Lett.* **51**, 1757 (1983).

21. Caffrey, A. J. *et al.* *Pap. presented at Muon-Catalysed Fusion Workshop*, Jackson, 7–8 June (1984).
22. Anderson, A. N. *et al.* *Pap. presented at Muon-Catalysed Fusion Workshop*, Jackson, 7–8 June (1984).
23. Jones, S. E. in *Atomic Physics 9* (eds Van Dyck, R. S. Jr & Fortson, E. N.) 99–115 (World Scientific, Singapore, 1984).
24. Cohen, J. S. & Leon, M. *Phys. Rev. Lett.* **55**, 52 (1985).
25. Menshihov, L. I. & Ponomarev, L. I. *Zh. eksp. teor. Fiz. Pis'ma Redakt.* (submitted).
26. Petrov, Yu. V. Preprint no. LNPI-1058 (1985).
27. Gershtein, S. S. *et al.*, *Soviet Phys. JETP* **53**, 872 (1981).
28. Cohen, J. S. & Leon, M. *Phys. Rev. A* **33**, 1437 (1986).
29. Jackson, J. D. in *Proc. Muon-Catalysed Fusion Workshop* (ed. Jones, S. E.) (1984).
30. Ceperley, D. & Alder, B. J. *Phys. Rev. A* **31**, 1999 (1985).
31. Mueller, B. & Rafelski, J. *Phys. Lett.* **164B**, 223 (1985).
32. Breunlich, W. H. *et al.* *Phys. Rev. Lett.* **53**, 1137 (1984).
33. Bleser, E. J. *et al.* *Phys. Rev.* **132**, 2679 (1963).
34. Kammel, P. *et al.* *Phys. Rev.* **A28**, 2611 (1983).
35. Kammel, P. *Lett. Nuovo Cim.* **43**, 349 (1985).
36. Ponomarev, L. I. *Atomkernenergie-Kerntechnik* **43**, 175 (1983).
37. Gershtein, S. S. *et al.* *Soviet Phys. JETP* **51**, 1053 (1980).
38. Kravtsov, A. V. *et al.* *JETP Lett.* **40**, 877 (1984).
39. Aristov, Yu. A. *et al.* *Yad. Fiz.* **33**, 1066 (1981) [*Soviet J. nucl. Phys.* **33**, 564 (1981)].
40. Kravtsov, A. V. *et al.* Preprint no. LNPI-998 (1984).
41. Bystritsky, V. M. *et al.* *Soviet Phys. JETP* **57**, 728 (1983).
42. Balin, D. V. *et al.* *Phys. Lett.* **141B**, 173 (1984).
43. Petrov, Yu. V. *Nature* **285**, 466 (1980); *Atomkernenergie-Kerntechnik* **46**, 25 (1985).
44. Petrov, Yu. V. & Shabelski, Yu. M. *Soviet J. nucl. Phys.* **30**, 66 (1974).
45. Takahashi, H. *et al.* *Atomkernenergie-Kerntechnik* **36**, 195 (1980).
46. Watts, K. D. in *Proc. Muon-Catalysed Fusion Workshop* (ed. Jones, S. E.) (1984).
47. Miller, R. C. & Krakowski, R. A. in *Proc. Muon-Catalysed Fusion Workshop* (ed. Jones, S. E.) (1984).
48. Chapline, G. Jr & Moir, R. Preprint no. UCRL-93611 (1986).
49. Jones, S. E. *Fusion Technol.* **8**, 1521 (1985); Preprint no. EGG-SE-6290, Idaho natn. Engng Lab. (1983).
50. Eliezer, S., Tajima, T. & Rosenbluth, M. N. Preprint no. DOE-ET-53088-223, Inst. Fusion Studies (1986).
51. Jackson, J. D. Preprint no. LBL-18266 (1984).
52. Schwarzschild, B. *Phys. Today* **39**, 18 (1986).
53. Zweig, G. *Science* **201**, 973 (1978).
54. Sakharov, A. D. & Lebedev, P. N. *Physics Inst. Rep.* (1948).
55. Petitjean, C. *Pap. presented at Am. Phys. Soc. Meet.*, Washington, DC, 29 April (1986).

ARTICLES

Splice-site selection by a self-splicing RNA of *Tetrahymena*

Richard B. Waring^{*§}, Paul Towner^{*}, Stephen J. Minter^{††} & R. Wayne Davies^{*††}

^{*} Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

^{††} Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario, Canada L4V 1P1

5' Splice-site selection by the self-splicing Tetrahymena ribosomal RNA precursor requires the formation of a particular RNA helix by base pairing between an intron 'guide sequence' and bases surrounding the 5' splice junction. Part of the guide sequence subsequently holds and positions the 3' end of the cut 5' exon to attack the 3' splice junction.

THE way in which splice sites are selected so precisely is a central question for all systems in which RNA splicing occurs, for a particular phosphodiester bond must be singled out from thousands of bonds that are chemically nearly identical. This specificity can be conferred only by higher-order molecular structure. The intron of the large ribosomal RNA precursor of *Tetrahymena thermophila* is excised, and the flanking exons ligated together in the absence of proteins^{1,2}. In this case, the splice junctions are defined, and catalysis carried out, by elements of RNA structure only. The capacity to form a particular set of RNA helices, including several involving distant RNA sequences, is conserved throughout the class of introns to which the *Tetrahymena* intron belongs³. Davies *et al.*³ showed that the capacity for exon sequences near the 5' and 3' splice junctions to base pair with an intron 'internal guide sequence' was also conserved, and proposed that these RNA helices could be important for defining the splice junctions and bringing both ends into the active site. We have used second-site suppression as a genetic tool to investigate the roles of these RNA structural elements in this RNA-catalysed RNA splicing reaction.

The RNA self-splicing reaction can be regarded as consisting of two subreactions, one at the 5' and one at the 3' splice junction—a simultaneous cleavage and ligation reaction in which phosphodiester bond energy is conserved—occurring via *trans*-esterification (phosphoester transfer)^{1,2}. The intron RNA takes up a very specific structure^{3–9}, binds a guanosine cofactor¹⁰ and can then catalyse the RNA splicing reaction if magnesium ions are present¹. The intron is excised as a unique linear

molecule with the guanosine added to the 5' end, and subsequently autocyclizes^{11–15}. All genetic^{16–20} and biochemical²¹ evidence is consistent with an RNA secondary structure closely related to that shown in Fig. 1A^{3,5}. The *Tetrahymena* ribosomal RNA intron is the biochemically best-studied member of a widespread class of introns (class I of fungal mitochondria) all of which are potentially self-splicing and have this RNA structure^{5,6,8}. Two mitochondrial introns of this class have been shown to be self-splicing^{22,23}.

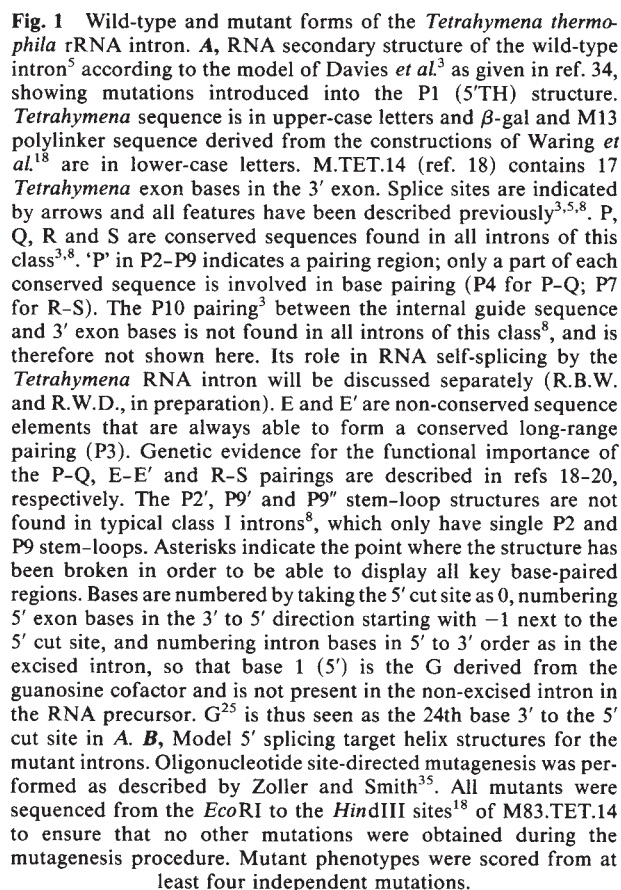
Each subreaction involves an attacking hydroxyl group and a target phosphodiester bond. We show that the formation of a short RNA helix involving an intron '5' guide sequence' and bases around the 5' splice junction is essential for 5' splicing. The structure of this helix and its position within the complete RNA structure define which phosphodiester bond is the 5' splice site. Part of the 5' guide sequence has a further role, positioning the attacking hydroxyl group in the 3' subreaction, which can be separated from the 5' subreaction. We are thus able to begin to dissect the RNA splicing reaction, and provide the first direct evidence that identifies elements of the mechanism by which a splice junction is selected.

Role of base pairing

A definitive way of demonstrating that interaction occurs between distant portions of RNA (or protein) molecules is to obtain second-site revertants of mutations that affect function. If, in an RNA molecule, the suppression of the original mutant phenotype by the second-site mutation is correlated with the restoration of the capacity to form a base pair disrupted by the first mutation, it follows that the capacity to form this base pair is important for the function that is altered in the mutant phenotype. The base-paired region may be a semi-permanent feature of the RNA structure, or may form transiently but still

† To whom correspondence should be addressed at Allelix Inc.

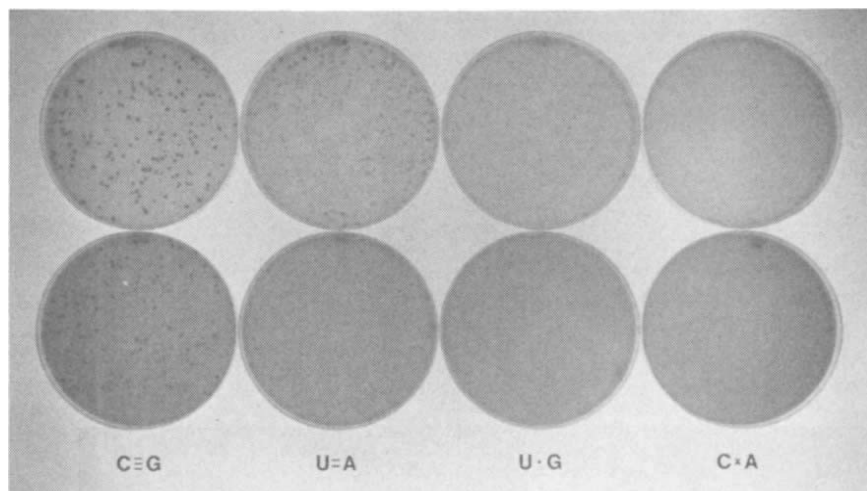
§ Present address: Department of Pharmacology, UMDNJ, Rutgers College of Medicine, PO Box 101, Piscataway, New Jersey 08854, USA.



We have made two sets of mutations in the short RNA helix named P1, which we now refer to as the 5' splicing target helix (5' TH) (see Fig. 1). Mutations affecting splicing were easily identified using the system of Waring *et al.*¹⁸. M83.TET.14 is an M13mp8.3 bacteriophage derivative, with the *Tetrahymena* intron (and short sections of the flanking exons) inserted so that the synthesis of the β -galactosidase α -fragment is prevented unless the intron excises itself *in vivo* in *Escherichia coli*. Phages

Oligonucleotide-directed *in vitro* mutagenesis was used to generate the intron mutation G²⁵A. The G²⁵A mutant itself gives a white plaque on minimal medium (Fig. 2) and is therefore more defective than the C⁻⁴U mutant. The double mutant C⁻⁴U, G²⁵A has a plaque colour phenotype slightly less blue than that of the wild type (Fig. 2), indicating that the C⁻⁴U mutation

Fig. 2 Plaque phenotypes of M83.TET.14 and M83.TET.14 carrying mutants in the 5' splicing target helix. The base pair present at position -4 in 5'TH (P1; Fig. 1) is given in each case to show the correlation between phenotype and the stability of 5'TH. From left to right, the phages are M83.TET.14 (C≡G), C⁻⁴UG²⁵A (U=A), C⁻⁴U (U·G) and G²⁵A (C/A). The top row of plates are of minimal medium, the lower row of YT medium. More intron excision is needed on YT than on minimal medium to produce the same plaque phenotype, which helps to distinguish between partial mutations¹⁸. For example, C⁻⁴U and C⁻⁴UG²⁵A are both a fairly similar pale blue on minimal medium, but C⁻⁴U is virtually white on YT and C⁻⁴UG²⁵A still pale blue.



suppresses the phenotype of the G²⁵A mutation and vice versa; the only reasonable explanation for this is that C⁻⁴ and G²⁵ are normally base paired, that C⁻⁴U and G²⁵A weaken or destroy this base pair, and that C⁻⁴U, G²⁵A replaces the original C·G base pair with a U·A base pair. The plaque phenotypes of these four phages, and therefore the efficiency of intron excision, are precisely in the order of the known strength of the base pair that each phage has in the -4 position of 5'TH—C≡G>U=A>U·G>C/A (Fig. 2)—as would be expected if base pairing underlies the function affected.

A second set of mutations (Fig. 1B) affecting different bases also predicted³ to be involved in forming 5'TH was made by *in vitro* mutagenesis, with two adjacent bases being altered for generating more extreme phenotypes, so as to facilitate *in vitro* RNA splicing analysis. The mutant phages are M83.TET.14 U⁻³A, C⁻²G (5THex; Fig. 1B) and G²³C, A²⁴U (5THint; Fig. 1B), both of which give white (intron excision-defective) plaques on minimal medium. The predicted second-site revertant combination (5THsup; Fig. 1B) would have a 5'TH containing the same base pairs as the wild type, in the same order, but with the purine-pyrimidine positions reversed (Fig. 1B). 5THsup plaques are almost as blue as wild-type plaques. Therefore, the pairs of double base-changes suppress each other's effects on intron excision. The occurrence of second-site suppression shows that any effects due to alteration of the β -galactosidase sequence by the exon mutants are not of primary importance, and that the capacity of the affected sequences to base pair is essential for normal intron function.

The phenotypes of these mutants were examined further in *in vitro* RNA splicing assays. Introns (with flanking exons as in Fig. 3A) carrying 5THex, 5THint or 5THsup mutations were subcloned into pSP65 (Fig. 3A) and transcribed *in vitro*; RNA was prepared and *in vitro* RNA splicing reactions were performed. Figure 3B shows the results of a typical *in vitro* RNA splicing experiment. The relevant data here are the levels of excised intron (IVS) and ligated exons (5E3E) produced; the origin of the novel products Y and Z is described below. The 5THex mutant is completely defective, producing no excised intron at all; clearly, this mutant is blocked in the very first step of RNA splicing. The 5THint mutant is slightly less defective, producing a very low but detectable amount of excised intron only in the presence of GTP; the GTP dependence of this reaction shows that the excised intron RNA is produced by the standard splicing reaction, and not by some side reaction. 5THsup produces a significant amount of excised intron in a GTP-dependent reaction, showing that the splicing defects of 5THex and 5THint are suppressed *in vitro*. The level of intron excised is not fully that of the wild type, and significant amounts of the novel products Y and Z generated by the 5THint mutant are still made by 5THsup, so that suppression is not total.

These results show that if we alter the capacity of a particular intron sequence—which we refer to as the 5' guide sequence (5'GS:5'UUUGGAGGG3'; Fig. 1A)—to base pair with the exon sequence adjacent to the 5' splice junction, RNA splicing is severely blocked. As all the base pairs affected are precisely those that would be centrally located in 5'TH (Fig. 1A), it follows that the capacity to form 5'TH is crucial for at least the 5' cleavage and ligation reaction. There is corroborative *in vivo* genetic evidence from yeast mitochondrial class I introns, as splicing-defective 5' exon mutations that would affect the 5'THs of cytochrome *b* introns 4 (refs 16, 24) and 5 (P. S. Perlman, personal communication) have been isolated, and one of the classes of revertants of the *cobI4* mutation is in 5'GS and would restore base pairing in 5'TH.

Factors influencing splicing rate

To determine the extent of phenotypic suppression in the 5THsup intron, a time course of the *in vitro* splicing reactions of wild-type and mutant introns (Fig. 4) was carried out. If the base pairing in 5'TH is restored, the splicing defect is suppressed, but splicing is still slow *in vitro*, reaching in 48 min less than half the level of excised intron produced by wild type in 1 min. This is not a result of depletion of the precursor by the reaction yielding Y and Z, as the generation of these products is particularly favoured in RNA transcripts generated from *Pvu*II-cut pSP65 (see below), and RNA splicing of 5THsup in transcripts from *Hind*III-cut pSP65, which produce very little Y, occurs at much the same rate (data not shown). The number of hydrogen bonds that can be formed between bases in the wild-type and 5THsup 5' splicing target helices is the same, so that it is the alteration in the sequence of base pairs in 5'TH that changes the rate of RNA splicing. Moreover, 5THint produced some excised intron and ligated exons but 5THex did not.

Separable 5' and 3' reactions

A striking property of the 5THint mutant is the production of large amounts of two completely new RNA species, bands Y and Z, in a GTP-independent reaction. Because the guanosine cofactor is required¹ for the nucleophilic attack that initiates the 5' cutting reaction, Y and Z must be produced by a reaction that bypasses 5' cutting, such as a novel version of the 3' cleavage-ligation reaction. Also, the function affected by the 5THint mutant is a function of the intron strand involved in the formation of 5'TH, and not of the exon strand of 5'TH or 5'TH itself. Both 5THint and 5THex mutations (Fig. 1A) disrupt the same two base pairs of 5'TH, yet 5THint makes large amounts of Y and Z, whereas wild-type and 5THex make no Z and very little Y (Fig. 3B). The structure of Y and Z reveals the nature of the bypass reaction. Y behaves as a linear molecule in different gel conditions, and is 437 nucleotides long, which corresponds

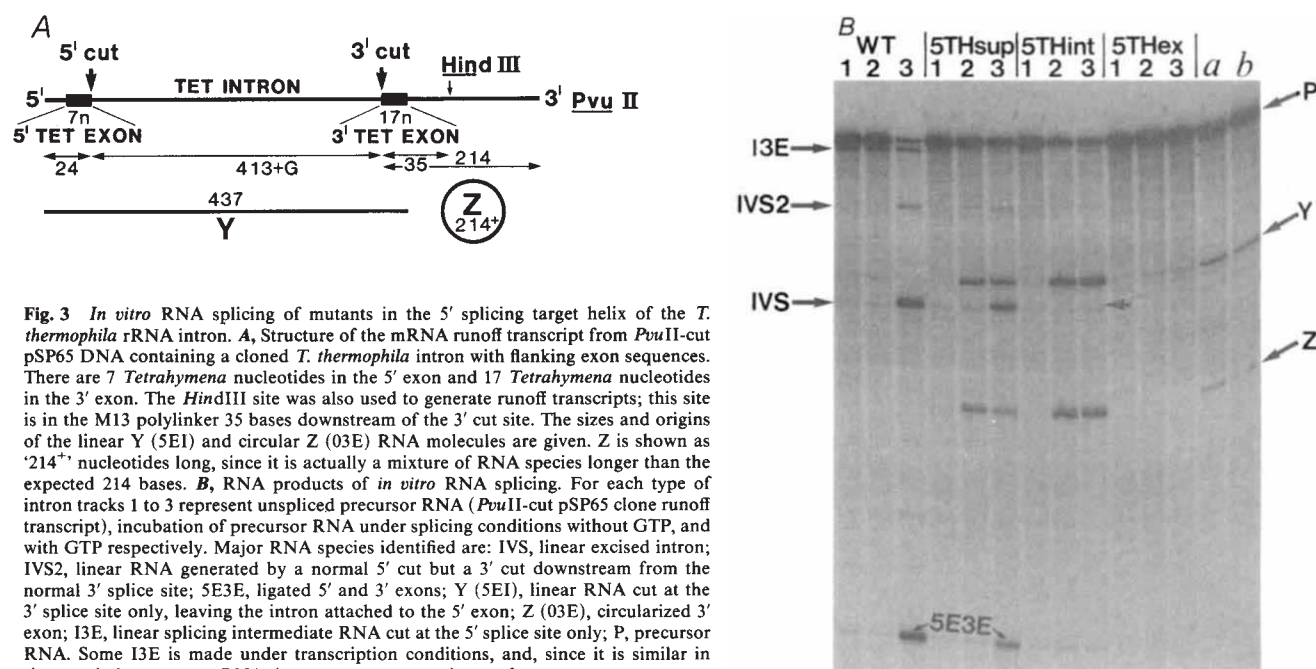


Fig. 3 *In vitro* RNA splicing of mutants in the 5' splicing target helix of the *T. thermophila* rRNA intron. **A**, Structure of the mRNA runoff transcript from *PvuII*-cut pSP65 DNA containing a cloned *T. thermophila* intron with flanking exon sequences. There are 7 *Tetrahymena* nucleotides in the 5' exon and 17 *Tetrahymena* nucleotides in the 3' exon. The *HindIII* site was also used to generate runoff transcripts; this site is in the M13 polylinker 35 bases downstream of the 3' cut site. The sizes and origins of the linear Y (5E1) and circular Z (03E) RNA molecules are given. Z is shown as '214' nucleotides long, since it is actually a mixture of RNA species longer than the expected 214 bases. **B**, RNA products of *in vitro* RNA splicing. For each type of intron tracks 1 to 3 represent unspliced precursor RNA (*PvuII*-cut pSP65 clone runoff transcript), incubation of precursor RNA under splicing conditions without GTP, and with GTP respectively. Major RNA species identified are: IVS, linear excised intron; IVS2, linear RNA generated by a normal 5' cut but a 3' cut downstream from the normal 3' splice site; 5E3E, ligated 5' and 3' exons; Y (5E1), linear RNA cut at the 3' splice site only, leaving the intron attached to the 5' exon; Z (03E), circularized 3' exon; I3E, linear splicing intermediate RNA cut at the 5' splice site only; P, precursor RNA. Some I3E is made under transcription conditions, and, since it is similar in size to whole precursor RNA, it occurs as a contaminant of some precursor preparations. Note that a minor band present at essentially the same place as 5E3E is of unknown origin, and is present in precursor preparations. The low levels of 5E1 (Y) which occur in wild type (WT) and the 5THex mutant probably result from the known tendency of this intron to undergo hydrolysis at the 3' splice junction at a low rate¹⁵. The two tracks on the far right are *in vitro* RNA splicing with precursors of mutant 29 (track a) which produces more 5E1/03E than wild type, but has a wild-type 5'TH (R.B.W. and R.W.D., unpublished), and the 5THex mutant (track b), incubated under splicing conditions with no GTP and 4 mM spermidine, showing stimulation of 5E1 and 03E production by spermidine, which can stabilize weak RNA helices^{26,27}.

Methods. The 458-bp *EcoRI*-*HindIII* fragments from M83.TET.14 mutants were purified and shuttled into the pSP65 transcription vector²⁵. pSP65 clones were linearized with *PvuII*, phenolized and ethanol-precipitated. RNA was synthesized by incubating 2 µg of DNA in the recommended buffer²⁵ including 4 mM spermidine (which promotes transcription, but partially inhibits splicing¹), bovine serum albumin (100 µg ml⁻¹), RNasin²⁵ and 5 U of SP6 RNA polymerase at 37 °C in a final volume of 50 µl. ³²P-GTP (20 µCi, 400 Ci mmol⁻¹; Amersham) in the presence of 20–50 µM GTP was used to label the RNA, which was then phenolized, ethanol-precipitated, resuspended in water and loaded onto a 7 M urea/5% polyacrylamide gel after the addition of at least 1 vol. of loading dye (deionized formamide containing 40 mM EDTA, 0.1% bromophenol blue). The gel was made and run exactly as a DNA sequencing gel³⁶. *HinfI*-digested M13mp8 was used as a size marker. RNA was located by autoradiography. Acrylamide strips were cut out, crushed and the RNA eluted overnight in 0.3 M Na-acetate, 20 mM Tris pH 8.0, 1 mM EDTA at 4 °C. Gel fragments were removed by centrifugation or filtration through siliconized glass wool and the RNA was then ethanol-precipitated twice at –20 °C. Attempts to clean RNA preparations using a variety of affinity chromatographic media, reverse-phase chromatography and Sephadex G-50 did not increase the initial rate of splicing and frequently reduced yields. RNA was stored frozen at –20 °C in 10 mM Tris pH 7.4, 0.1 mM EDTA. Unless otherwise indicated, splicing reactions¹ were performed in siliconized 0.5-ml Eppendorf tubes in 10 mM (NH₄)₂SO₄, 30 mM Tris pH 7.5, 5 mM MgCl₂ and 0.2 mM GTP at 37 °C for 30 min. Splicing was stopped by addition of at least 1 vol. formamide loading dye³⁶. Samples were run on polyacrylamide gels as described above. Gels were fixed and usually exposed at room temperature without an intensifier screen.

precisely to the size of molecule consisting of the joined 5' exon and intron, '5E1'. The presence of the 5' exon in Y was confirmed by generating a complementary DNA runoff transcript of the expected size using an oligonucleotide that primes only 18 bases downstream of the 5' splice junction within the intron. The mobility of Z varies with the strength of the polyacrylamide gels, as is typical of circular molecules. Purified Z, when nicked, gives two linear molecules 215 and 216 nucleotides long in roughly equal proportions, corresponding to the expected length (Fig. 3A) of the 3' exon plus one or two nucleotides respectively. As sequencing (Fig. 5) established that the 3' end of the precursor transcript had been joined to the first base of the 3' exon, Z must be the circularized 3' exon, '03E'. 03E contains either one or two extra nucleotides at the circularization point which are not in the DNA template. The runoff transcript itself (and all derivatives of it which contain the 3' exon) is an approximately equal mixture of RNA molecules with one or two additional nucleotides which must be at the 3' end and are presumably added spontaneously by SP6 polymerase, an effect observed previously²⁵. Thus, the bypass reaction is a cleavage and ligation reaction involving cleavage at the normal 3' splice site, with ligation of the free 3' end of the RNA precursor to the 5' end of the cut 3' exon, producing a circular 3' exon (Z, 03E) and a linear molecule comprising the intron joined to the 5' exon (Y, 5E1).

The generation of 5E1 and 03E by the 5THint mutant is 1/10th to 1/20th as rapid as normal splicing by the wild-type intron, and thus 5THint RNA carries out a rapid 3' splicing reaction

analogue while it is almost completely blocked for 5' cleavage and ligation. 5E1 molecules, which lack the 3' exon, also undergo 5' cleavage in the absence of a 3' reaction, and the relative rates of 5' cleavage for molecules with wild-type, 5THint, 5THex and 5THsup sequences are as expected from the above results (data not shown). The 5' and 3' cleavage and ligation reactions are separable, and 5'TH is involved only in the 5' subreaction.

Positioning of attacking group

The fact that the high 03E and 5E1 production results from the 5THint mutation means that the intron sequence altered by this mutation must be involved. In the runoff transcript from *PvuII*-cut template, the RNA precursor should end in 5'CCAG3' (actually it ends with 5' CCAGN(N')3'; Fig. 5), which can pair precisely with the sequence 5'CUGG3' containing the double base change (GA→CU) of 5THint compared with wild type. The 3' end of the transcript can then substitute for the cut 5' exon. This hypothesis predicts that the phenomenon will be specific to DNA transcripts containing a *PvuII* end, since sequences adjacent to other restriction enzyme sites will not pair as well with the 5'TH intron strand of 5THint. Table 1 shows that this is the case; the reaction generating 5E1 and 03E is much less frequent using runoff transcripts from templates containing *HindIII*, *HpaII* or *HinfI* ends. The amounts of 5E1 and 03E produced actually correlate well with the strength of potential pairing between the end of the RNA transcript and the key region of intron RNA (Table 1), and their amounts are increased

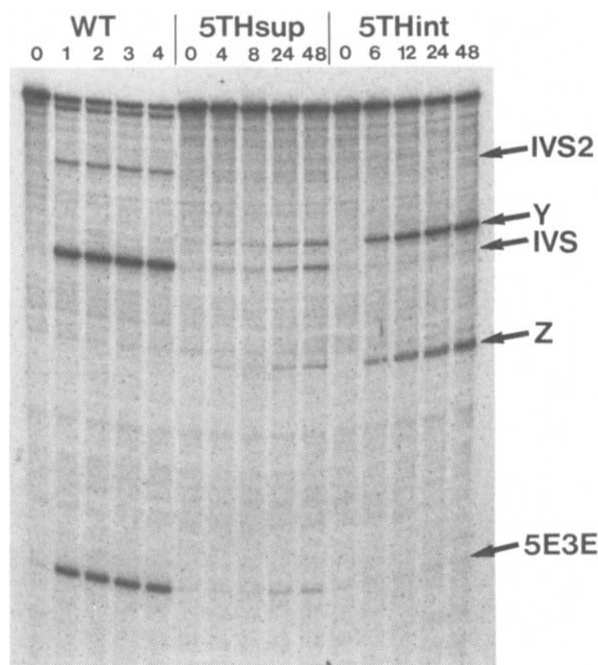


Fig. 4 Time course of *in vitro* RNA splicing of mutant and wild-type *Tetrahymena* rRNA introns. Numbers above each track indicate the time in minutes at which samples were taken. Runoff transcripts from *Pvu*II-cut pSP65 clones were generated, purified and incubated under splicing conditions as described for Fig. 3B, except that splicing was performed at 30 °C, rather than 37 °C. Major RNA species are labelled as in Fig. 3B.

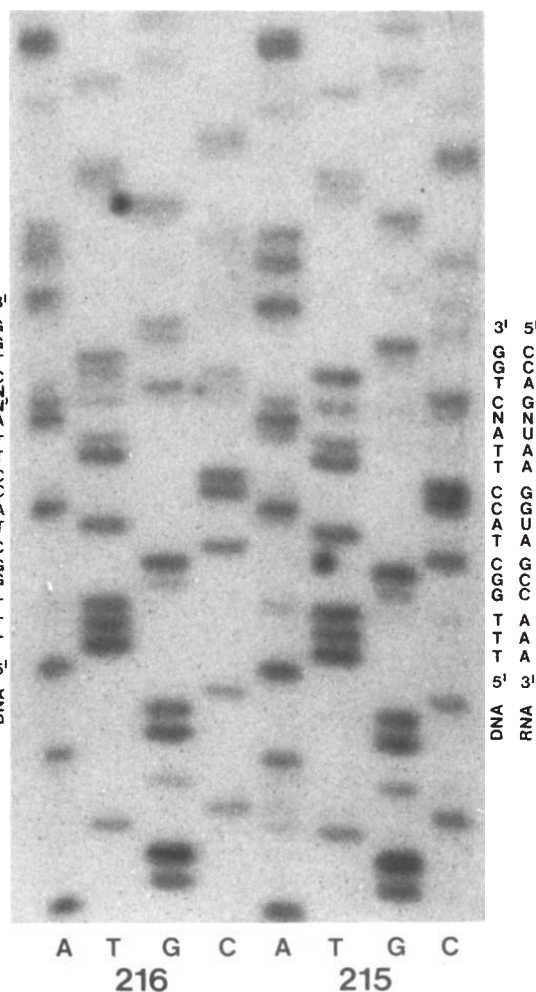


Fig. 5 Nucleotide sequencing of the circularization site of the circular 3' exon (03E). The two 03E bands (215 and 216 nucleotides long) were separated, and the sequences of both are shown side by side; note that the sequences run parallel until the circularization site, where they shift 1 base out of phase. DNA sequences read directly from the gels are shown alongside, together with the RNA complement which is the true sequence of the 03E molecule in this region. The sequence appears to become unclear after the circularization point; this is because the two 03E species could not be obtained completely without cross-contamination, and their sequences are, of course, shifted by 1 base relative to each other after the circularization point. The 215 03E has one extra base at the circularization site, the 216 03E has two. In the RNA, N is A>U>G>C in the 215-nucleotide 03E. In the 216-nucleotide 03E, N is C>A, U, G, N' is U>A>C>G.

Methods. RNA was purified as described for Fig. 3. Sequencing was carried out using the dideoxy procedure with reverse transcriptase at 42 °C and priming from ³²P-end-labelled M13 reverse sequencing primer (Amersham), which pairs with the downstream exon. The reaction mixes were treated with RNase before boiling and loading onto gels. 03E RNA was produced in sufficient amounts for this experiment by incubating precursor transcript from a pSP65 clone of the 5THint mutant intron under splicing conditions for 1 h. More than 90% of the precursor was converted to 5EI and 03E in this time.

in the presence of spermidine, which stabilizes weak RNA helices^{26,27}.

Our experiments identify the binding site within the catalytic RNA molecule for the nucleotide sequence (normally the 3' end of the cut 5' exon) that carries the 3' hydroxyl attacking group for the second (3') step in the RNA splicing reaction; we will refer to the binding site as the 3' attack site (3'AS). The nucleotides making up 3'AS are a subset of those in the 5' guide sequence. In the 5THint mutant the sequence of 3'AS is 5'(G)CUGG3'; the wild-type equivalent is 5'(G)GAGG3'. Figure 6 outlines the role of 3'AS in the generation of 5EI and 03E. The reaction described here ('3' end invasion') is intramolecular, but a clear prediction is that any RNA fragment or oligonucleotide that can pair with the bases 5'[G]GAGG[G]3' to place a 3' hydroxyl end correctly will be used as an alternative to the cut 5' exon if provided in sufficient concentration ('oligonucleotide invasion'). This is consistent with the recent results of Cech and co-workers²⁸. Mutants with a weak, unstable 5'TH will be particularly susceptible to oligonucleotide invasion, which would lead to *trans* activity of this RNA catalyst.

Conclusions

Our results lead us to draw three major conclusions about the mechanism of RNA splicing for self-splicing class I introns: (1) the formation, at least transiently, of a short RNA helix (5'TH; 5' splicing target helix) between an intron 5' guide sequence and bases surrounding the 5' splice junction, is essential for the selection of the 5' splice site and for 5' cleavage and ligation; (2) the 5' and 3' cleavage and ligation reactions are completely separable; (3) a subset of the intron bases in the guide sequence (those that pair with 5' exon bases) serves to retain and position the cut 5' exon, or any externally supplied partner sequence with a free 3' OH end, to attack the 3' splice junction. In Fig. 6 these conclusions are incorporated into the present understanding of the RNA-catalysed splicing reaction.

As the 3' splicing bypass reaction occurs at a significant rate (1/10th that of normal splicing), some detailed inferences concerning the structures involved can be drawn. First, the nature of the last base of the cut 5' exon is not crucial for 3' cutting, as this base is highly variable on the 3' end of the precursor transcript due to the semi-random addition of nucleotides by SP6 RNA polymerase. Also, the hydroxyl that attacks the 3'

Table 1 Alternative base pairing by an intron sequence involved in structures essential for 5' or 3' cleavage and ligation

	5'TH	3'AS/RNA 3' End	5EI(Y)	03E(Z)	I/5E3E
	5'TH 5'ACUCUAAA ::: : AGGUGUUU	PvuII 5'UGCCAGN(N) : : : : AGGUGUUU	++++(+ + + + +)	++++(+ + + + +)	+
	5'THsup 5'ACUCUAAA : : : : : AGGUGUUU	PvuII 5'UGCCAGN(N) : : : : AGGUGUUU	+++(+ + +)	+++(+ + +)	+++
	wt 5'ACUCUAAA : : : : : AGGAGGUUU	PvuII 5'UGCCAGN(N) : : : AGGAGGUUU	+(+)	-(+)	++++
	5'THex 5'ACUCUAAA : : : : : AGGAGGUUU	PvuII 5'UGCCAGN(N) : : : AGGAGGUUU	+(+)	-(+)	-
		HpaII AGCCGN(N) : : : AGGUGUUU	+(+)	-(+)	+
		HinfI 5'TH UUAUGAUN(N) : : : : AGGUGUUU	+(+)	-(-)	+

The potential to form 5'TH or 3'AS-RNA 3' end structures is correlated with the production of normal splicing products (I/5E3E) and unusual RNA molecules derived by 3' cleavage and ligation only (5EI, Y; 03E, Z) in *in vitro* RNA splicing reactions. 5'TH, 5' splicing target helix (P1 in refs 3, 5, 8, 9); 3'AS, 3' attack site; 5EI, linear fused 5' exon and intron; 03E, circular 3' exon; I, excised intron; 5E3E, ligated 5' and 3' exons. 5'TH is essential for 5' cleavage-ligation, and the 3'AS-RNA 3' end pairing is proposed as the basis of the 3' end invasion reaction that generates 5EI and 03E. The strongest base pairing possible is given in all cases, whether or not such a pairing is potentially stable. The arrow in 5'TH indicates the 5' splice junction. The 3' end of the RNA transcript is shown with one or two semi-randomly added nucleotides, as described in Fig. 5 legend. 5'TH, 5'THex and 5'THsup mutants are described in the text; bases differing from wild type are underlined. PvuII, HpaII and HinfI refer to the restriction enzymes used to cut the pSP65 clone DNA template, which determines the end of the runoff transcript. The lengths of the 3' exons, not including bases added by SP6 polymerase, are: HindIII, 35; HpaII, 110; PvuII, 214; HinfI, 228. Data for transcripts from HindIII-cut DNA are not included here, as the very small 03E from this reaction was not identified clearly. Production of various RNA molecules is scored semi-quantitatively on a scale of 0(-) to 5(+++++), based on data obtained under standard splicing conditions. Symbols in parentheses refer to data obtained in the presence of 4 mM spermidine under normal splicing conditions. The production of 03E is the crucial indicator of the 3' end invasion reaction, as low background levels of 5EI are produced by spontaneous scission at the 3' splice site¹⁵. The relative success of the normal 5' splicing and 3' end invasion reaction depends on: (1) the potential for base-pairing between the 3' end of the transcript and 3'AS; (2) the relative stability of 5'TH, the 3'AS-5' exon pairing, and the 3'AS-RNA 3' end pairing. With 5'TH in a transcript from PvuII-cut DNA, normal splicing is rare because of the defective 5'TH, but the bypass reaction is frequent because of the pairing of the 5'CU3' sequence with the 3' end of the RNA transcript. In wild type, 5'TH is strong and the alternative pairing with the end of the transcript is weak so that the bypass reaction rarely if ever occurs; in 5T-ex, which has a wild-type intron strand, 03E production is detectable when the weak alternative pairing is stabilized by spermidine (Fig. 3B). 5'THsup can form as good a pairing with the transcript's 3' end as 5'TH, but has a much stronger 5'TH, so that 5EI/03E generation is reduced and normal splicing increased in 5'THsup over 5'TH. Note that the presentation of the 3' attacking group is achieved adequately by two very different short helices in the wild-type/cut 5' exon and 5'TH/RNA 3' end combinations. In the role of 3'AS the 5'GGAGGG3' sequence primarily determines the position of the 3' hydroxyl group, placing it in close proximity to its target, the 3' splice junction, and is thus truly acting as an internal guide sequence as proposed by Davies *et al.*³.

splice junction can be at least one base beyond the usual position, since the '+2' transcript directs 3' splicing. Finally, the U carrying the hydroxyl that attacks the 3' splice junction in the 'normal pathway' (Fig. 6) need not be paired with G as in 5'TH because the '+1' structure has a 3'A^{OH} that is unpaired yet gives good 3' splicing.

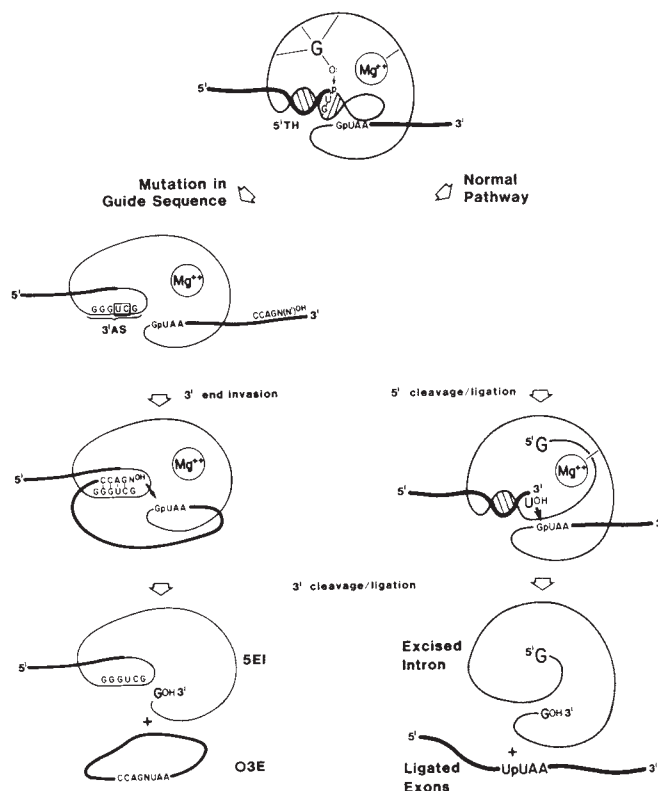


Fig. 6 Diagram of the steps in the RNA splicing reaction catalysed by the intron of the rRNA precursor of *T. thermophila*, and the events leading to the generation of 5EI and 03E by the 3' end invasion reaction in mutants that alter the sequence of the 5' guide sequence, such as 5'TH. The boxed bases are the mutated bases in the guide sequence of the 5'TH mutant, 5'CU3' replacing the 5'GA3' of wild type. The RNA precursor shown is an *in vitro* transcript from PvuII-cut DNA of a clone of the 5'TH mutant *T. thermophila* rRNA intron. Exon RNA is shown as a thick line, intron RNA as a thin line. 3'AS, 3' attack site; 5'TH, 5' splicing target helix; G, guanosine cofactor. The figure is based on conclusions from the present paper and the experiments of T. R. Cech and co-workers^{1,2,10,15}.

The formation of 5'TH is necessary but not sufficient for effective 5' splicing, since particular series of base pairs in this helix allow much more rapid reaction than others. It seems unlikely that the difference in stacking energy between the wild-type 5'TH and the 5'THsup 5'TH is actually significant enough to account for the difference in behaviour of 5'THsup and wild type in terms of stability of 5'TH. The sequence of 5'TH may be a factor in positioning the helix via tertiary interactions between exposed groups on bases in 5'TH and other parts of the RNA catalyst, and/or in determining the susceptibility to cleavage of the phosphodiester bond at the 5' splice site. Recent work on the crystal structure of A-form DNA helices containing T·G base pairs²⁹ suggests that U·G base pairs may introduce novel features into RNA helices, changing the availability of groups for tertiary interaction and altering local stacking. A U·G base pair is found in 5'TH next to the 5' splice site in all but one class I intron⁸. The purine-pyrimidine asymmetry of the *Tetrahymena* 5'TH, which switches strands precisely at the 5' splice site, may also be significant for this intron, but is not a general feature of class I introns. Either of these features could influence the lability of the bond at the 5' splice junction, and would be affected by local alterations in the structure of 5'TH.

Pairing of the 3' end of the RNA precursor with the 3'AS in 5'TH would be regarded as unstable on the basis of the rules generally used to establish the stability of RNA structures³⁰.

Also, in some introns of this class, noticeably the third and fourth introns of the yeast cytochrome oxidase gene and the fourth intron of the yeast apocytochrome *b* gene, 5'TH is very short^{3,8}, consisting of 4,3 and 4 base pairs (bp), respectively. A role for protein factors such as maturases³¹ in stabilizing or directing the folding of RNA can be envisaged; indeed, the intron specificity of maturases could be explained if they stabilize guide sequence-exon interactions rather than other parts of the molecule where conserved sequences underlie the structure. In addition, the RNA structure must be viewed as being dynamic rather than static. The 5THsup mutant RNA is a clear example of this; during the long interval that elapses before a significant proportion of the molecules are cut at the 5' splice junction (due to the aberrant nucleotide sequence in 5'TH), the 3' end of the RNA transcript is able to find the intron strand of 5'TH unpaired often enough to generate a considerable level of the 3' cleavage bypass reaction. Interactions involving only 3 bp have been shown to result in 'reverse cyclization' of the excised intron³², and CpU can act as a 5' exon²⁸, presumably by pairing with the 5' end GG of 3'AS. In a dynamic, rather than a static, RNA structure, the concept of stability must be used more flexibly, and the establishment of classical 'stable pairings' should not necessarily be used as a prerequisite when seeking to determine the structural elements important in dynamic processes such as

RNA splicing or ribosome function.

We have presented here the first direct evidence concerning the mechanism of splice-site selection in a ribosomal or messenger RNA precursor, identifying key elements of RNA structure involved in 5' and 3' cleavage and ligation. These elements, within the context of the whole RNA molecule, provide the information that defines particular phosphodiester bonds as the targets for cleavage and ligation. The sequence of 5'TH and its position relative to the attacking guanosine 3' hydroxyl and the Mg²⁺ ion determines the 5' splice junction. The 3'AS/cut 5' exon helix positions the attacking hydroxyl group at the end of the 5' exon correctly relative to the Mg²⁺ ion and the 3' splice junction. Similar elements of RNA structure may define splice junctions in nuclear mRNA precursors. An intact 5' end of the RNA moiety of U1snRNA is essential for splicing³³, and this RNA sequence can pair with bases near 5' splice junctions to produce a structure similar to 5'TH. Precise understanding of splice-site selection requires crystal structures of 5'TH and the whole RNA.

This work was supported by a grant from the MRC to R.W.D. and Dr C. Scazzocchio. We thank Heather Erfle of Allelix Inc. for assistance in the synthesis of oligodeoxyribonucleotides, and Dr C. Scazzocchio for his contribution to the initiation of this work, as well as for constant support and helpful discussion.

Received 5 December 1985; accepted 18 March 1986.

- Kruger, K. *et al.* *Cell* **31**, 147-157 (1982).
- Cech, T. R., Zaug, A. J. & Grabowski, P. J. *Cell* **27**, 487-496 (1981).
- Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. *Nature* **300**, 719-724 (1982).
- Michel, F., Jacquier, A. S. & Dujon, B. *Biochemie* **64**, 867-881 (1982).
- Waring, R. B., Scazzocchio, C., Brown, T. A. & Davies, R. W. *J. molec. Biol.* **167**, 595-605 (1983).
- Michel, F. & Dujon, B. *EMBO J.* **2**, 33-38 (1983).
- Cech, T. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 3903-3907 (1983).
- Waring, R. B. & Davies, R. W. *Gene* **28**, 227-291 (1984).
- Davies, R. W. *Biosci. Rep.* **4**, 707-727 (1984).
- Bass, B. L. & Cech, T. R. *Nature* **308**, 820-826 (1984).
- Grabowski, P. J., Zaug, A. J. & Cech, T. R. *Cell* **23**, 467-476 (1981).
- Brehm, S. L. & Cech, T. R. *Biochemistry* **22**, 2390-2397 (1983).
- Zaug, A. J., Grabowski, P. J. & Cech, T. R. *Nature* **301**, 578-583 (1983).
- Zaug, A. J., Kent, J. R. & Cech, T. R. *Science* **224**, 574-578 (1984).
- Zaug, A. J., Kent, J. R. & Cech, T. R. *Biochemistry* **24**, 6211-6218 (1985).
- De La Salle, H., Jacq, C. & Slonimski, P. P. *Cell* **28**, 721-732 (1982).
- Netter, P., Jacq, C. & Slonimski, P. P. *Cell* **28**, 733-738 (1982).
- Waring, R. B., Ray, J. A., Edwards, S. W., Scazzocchio, C. & Davies, R. W. *Cell* **40**, 371-380 (1985).

- Weiss-Brummer, B., Holl, J., Schweyen, R. J., Rödel, G. & Kaudewitz, F. *Cell* **33**, 195-202 (1983).
- Holl, J., Rödel, G. & Schweyen, R. J. *EMBO J.* **4**, 2081-2085 (1985).
- Inoue, T. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 648-652 (1985).
- Garriga, G. & Lambowitz, A. M. *Cell* **38**, 631-641 (1984).
- Van der Horst, G. & Tabak, H. F. *Cell* **40**, 759-766 (1985).
- Perea, J. & Jacq, C. *EMBO J.* **12**, 3281-3288 (1985).
- Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035-7056 (1984).
- Holbrook, S. R. *et al.* *J. molec. Biol.* **123**, 631-642 (1978).
- Quigley, G. J., Teeter, M. M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 64-70 (1978).
- Inoue, T., Sullivan, F. X. & Cech, T. R. *Cell* **43**, 431-437 (1985).
- Brown, T., Kennard, O., Kneale, G. & Rabinovitch, D. *Nature* **315**, 604-606 (1985).
- Tinoco, I. Jr, Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362-367 (1971).
- Jacq, C. *et al.* in *Mitochondrial Genes* (eds Slonimski, P. P., Borst, P. & Attardi, G.) 155-183 (Cold Spring Harbor Laboratory, New York, 1982).
- Sullivan, F. X. & Cech, T. R. *Cell* **42**, 639-648 (1985).
- Krämer, A., Keller, W., Appel, B. & Luhrmann, R. *Cell* **38**, 299-307 (1984).
- Davies, R. W., Waring, R. B., Brown, T. A. & Scazzocchio, C. in *Mitochondria 1983 Nucleo-Mitochondrial Interactions* (eds Schweyen, R. J., Wolf, K. & Kaudewitz, F.) 179-189 (de Gruyter, Berlin, 1983).
- Zoller, M. J. & Smith, M. *DNA* **3**, 479-488 (1984).
- Davies, R. W. in *Gel Electrophoresis of Nucleic Acids: A Practical Approach* (eds Rickwood, D. & Hames, B. D.) 117-172 (IRL Press, Washington, DC, 1982).

LETTERS TO NATURE

High-resolution imaging from Mauna Kea: the triple quasar in 0.3-arc s seeing

J. Patrick Henry & J. N. Heasley

University of Hawaii at Manoa, Institute for Astronomy,
2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

The 'triple' quasar, PG1115+08 (which actually contains at least four objects: A₁, A₂, B and C) was the second gravitational lens object to be found¹. But because of its small angular scale (<2.5 arc s separation between components) and lack of detectable radio emission, which precludes the use of high-resolution interferometric radio techniques, not much is known about this object. In particular, the lensing object(s) have not been identified. Here we report imaging observations made on a night of exceptional seeing, as well as spectroscopy of two galaxies near the quasar. We find that there is a galaxy centred approximately midway between the two A components. The properties of this galaxy are consistent with it being the brightest member of a small group at a group redshift of 0.305. Although detailed modelling is required, it is likely that the galaxy and its group are the lenses.

Previous imaging of PG1115+08 has been either by direct photography or speckle interferometry²⁻⁶; in all cases com-

ponents B and C have been unresolved. Component A is resolved into two subcomponents (termed A₁ and A₂, north and south, respectively) separated by ~0.5 arc s at a position angle of ~20°. The intensity ratio is ~1. All observations have failed to detect a lens galaxy, which must be fainter than V = 22 (visual magnitude) or be ≤1 arc s from a component. Furthermore, there is no cluster in the field brighter than a Thuan-Gunn *r* magnitude of 25 (ref. 2). The only objects previously known to be near PG1115+08 are three galaxies 12-24 arc s to the south-west, whose redshifts have not been measured until now.

We made direct imaging observations of PG1115+08 on 9 April 1983 UT (Universal Time) at an airmass of between 1.02 and 1.07. These observations were obtained with an RCA C33063 image intensifier mounted behind a Barlow lens and a Schott GG 475 filter at the Cassegrain focus of the University of Hawaii 2.2-m telescope on Mauna Kea. This combination yielded a bandpass of ~4,750-6,000 Å and a scale of 4.1 arc s per mm. The data were recorded on a roll of Ila-O film which was mounted in a camera capable of taking many sequential exposures of specified length while automatically advancing the film. From a previous run we determined that 2 s was the minimum exposure time that would record the faintest (B) component. We obtained 285 2-s exposures, 105 3-s exposures and 230 10-s exposures. The sky was well exposed on the 10-s exposures. Only the 2-s exposures were used for imaging because the seeing was noticeably worse on the 3-s exposures.

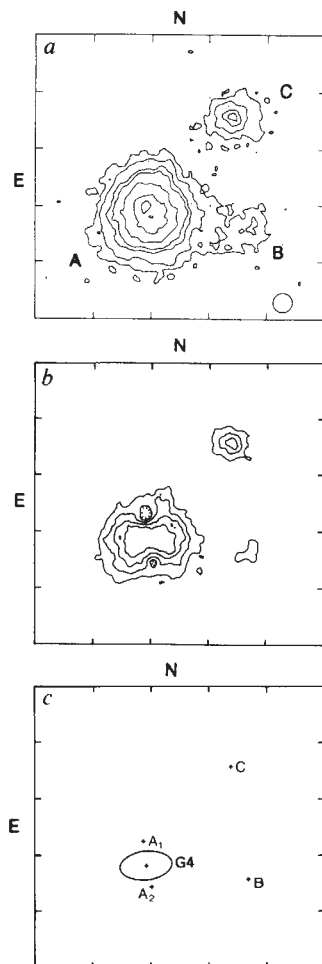


Fig. 1 Diagrams of the PG1115+08 field. The tick marks along the axes are at 1 arc s intervals. *a*, Contours of surface brightness in the 9 April 1983 summed image (see text for details). The contours are at 0.9, 0.8, 0.6, 0.4, 0.3, 0.15 and 0.07 of the peak. If component C did not vary during the one month between our observation and that of ref. 4 (our crude photometry verifies this assumption to 0.2 mag), then the peak surface brightness is $\mu_V = 16.8 \text{ mag (arc s)}^{-2}$. The resolution of this image is shown by the circle at the lower right which indicates the FWHM; it is slightly larger than that of the raw data because of the light smoothing to attenuate the high-frequency noise. *b*, Contour diagram after subtracting the best-fitting point sources in the two point source+galaxy model. The contours are at 0.8, 0.6, 0.4 and 0.2 of the galaxy brightness peak, which, under the assumptions given in *a*, is $\mu_V = 17.1 \text{ mag (arc s)}^{-2}$. Contours with tick marks indicate local minima in the surface brightness at the locations of the imperfectly subtracted point sources. *c*, Schematic diagram of the objects in the PG1115+08 field. The outline of the galaxy (G4) is at one core radius.

We examined all the exposures and selected the best 19 for further analysis. These images were scanned on a PDS 1010A microdensitometer; the aperture used was $12 \times 12 \mu\text{m}$, scanned in $12\text{-}\mu\text{m}$ steps; that is, a pixel size of 0.049 arc s . The maximum density anywhere in the field was only 1.2 density units. The digitized images were then converted to intensities using a calibration curve derived from a step exposure through the image tube onto the same roll of film, developed at the same time as the imaging data. We determined the full width at half maximum (FWHM) of component C in each of these 19 images. The best five were co-aligned and summed to produce our final data set.

During times of extremely good seeing at Mauna Kea, the image consists of a tiny spot dancing within a circle of diameter

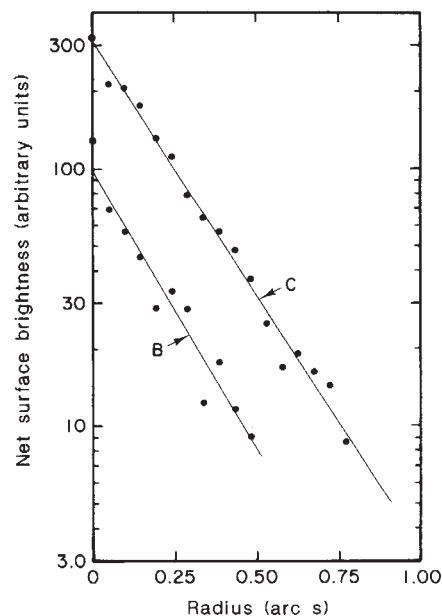


Fig. 2 Azimuthally averaged surface brightness of the B and C components.

$\sim 1 \text{ arc s}$; the above procedure was designed to remove the image jitter, leaving only the tiny spot. We show a contour diagram of the final image (slightly smoothed) in Fig. 1*a*.

Azimuthally-averaged profiles for components B and C are shown in Fig. 2. Fitting an exponential to the data yields a point spread function proportional to $\exp(-\theta/0.23 \text{ arc s})$ (C) and $\exp(-\theta/0.21 \text{ arc s})$ (B), implying FWHMs of 0.31 and 0.29 arc s, respectively. In what follows we adopt the point spread function determined from C because it has a higher signal-to-noise ratio. As the two point spread functions are virtually the same, it seems reasonable to assume that the same function will also be applicable to the region of component A. Although most previous descriptions of the seeing function have used a gaussian distribution, at least at small radii (see, for example, ref. 7), such a function does not fit our data, probably because we have eliminated the image motion. Integrating over such motion probably produces a gaussian point spread function by way of the central limit theorem.

The parameters of a nonlinear least-squares fit to a two-dimensional array of data around components B and C yielded the results given in Table 1, which also includes the results of previous measurements. The only significant difference is in the intensity (flux) ratios; in particular, there is a change in the ratio f_B/f_A on a timescale of a month between our data and those of ref. 4. Crude photometry, using the sky as a zero point, indicates that the B component weakened by $\sim 0.5 \text{ mag}$ during that time.

Figure 1*a* shows that A must be more complicated than the two equal-intensity point sources separated by 0.5 arc s detected previously. The intensity is approximately constant in an area 0.5 arc s in diameter at the peak, yet we should easily resolve the point sources if they were the only contributors to A. To quantify these statements, and also to better understand the structure of the A component, we have fitted a variety of models to our data. Our intensity data are unlikely to have a Poisson distribution, because of the nonlinearity of the photographic process; therefore, the usual goodness-of-fit tests cannot be used. Yet we need to know when to stop adding free parameters to the models. We adopt the following criterion, which seems reasonable, but is certainly not rigorous: A model for the A component will be considered acceptable if the root-mean-

Table 1 Parameters of PG1115+08B, C

Date	Sep(AB)* (arc s)	PA(AB)* (° E of N)	$f_B/f_A^\dagger$	Sep(AC)* (arc s)	PA(AC)* (° E of N)	$f_C/f_A^\dagger$	Ref.
9 April 1983	1.74 (5)‡	263 (1)	0.074 (5)	2.16 (5)	323 (1)	0.171 (6)	This work
7 March 1983	1.77 (2)	266 (2)	0.12 (2)	2.29 (3)	322.3 (8)	0.19 (2)	4
30 April 1981			0.12 (2)			0.17 (2)	4
17 June 1980	1.77 (3)	265.8 (7)	0.110 (6)	2.28 (2)	322.3 (4)	0.180 (4)	2
3 June 1980	1.9 (1)	263 (5)	0.113 (2)	2.4 (1)	317 (5)	0.174 (8)	3

* Separation (Sep) and position angle (PA) calibrated from those of star S given in ref. 2.

† The flux for component A (f_A) comes from the best-fitting two point+galaxy model described below. Other models change its value by about $\pm 10\%$.

‡ Number in parentheses is the error in the last digit.

Table 2 Parameters of PG1115+08A

Model	Sep(A ₁ A ₂) (arc s)	PA(A ₁ A ₂) (° E of N)	f_{A1}/f_{A2}	a (arc s)	ϵ	PA(G4)* (° E of N)	Sep(A ₁ G4) (arc s)	PA(A ₁ G4) (° E of N)	f_{G4}/f_A	Residual (r.m.s.)
2 points	0.57 (5)†	20 (1)	1.04 (2)	—	—	—	—	—	—	84
2 points+galaxy	0.80 (5)	10 (1)	0.85 (5)	0.47 (1)	0.38 (2)	97 (1)	0.43 (5)	193 (1)	0.46 (2)	56

* The position angle of the galaxy (G4) is that of its semimajor axis.

† Number in parentheses is the error in the last digit.

square (r.m.s.) sum of residuals from it is of the order of the mean of the fluctuations in the A peak (72) and the background (30); that is, 50.

Table 2 contains the results of fits of two models to the A component data. As expected, a model containing just two point sources has unacceptably large residuals, although the best-fitting parameters are virtually identical to those given in refs 4–6. We then tried a model with two point sources and a galaxy. The galaxy was modelled as having an elliptical surface brightness of major axis a , and ellipticity $\epsilon = 1 - (\text{minor axis}/\text{major axis})$. The surface brightness was assumed to fall off as $(1 + r^2/a^2)^{-1}$ along the major axis and correspondingly at other position angles. This form accurately approximates a King⁸ model in the core. Since the integrated surface brightness of this model diverges logarithmically, we give the light within the core when describing the galaxy brightness. Best-fitting parameters are given in Table 2. The residuals to this second model are acceptably small, although they do exhibit some systematic trends, such as the negative-going contours at the location of A₁ and A₂, shown in Fig. 1b (tick marks). A schematic diagram of the various objects in the field is shown in Fig. 1c, where the galaxy is labelled 'G4'.

Several tests can be made of the existence of G4. First, do other simple models substantially reduce the residuals? We have tried four additional models: three point sources, three point sources and a galaxy, and two 2 point+galaxy models with broader point spread functions (0.4 and 0.5 arc s). The lowest r.m.s. residual from these four models was 50 (3 points+galaxy). This reduction in the residuals is not large compared with that achieved by going from 2 points to 2 point+galaxy models, so we consider the latter to be a reasonable representation of the data. Second, will a broader point spread function eliminate the galaxy? We find that in order to decrease the galaxy intensity to near zero, it is necessary to increase the FWHM of the point response to 0.5 arc s, which certainly does not fit the data of B and C (see Fig. 2). Third, with a galaxy as bright as we claim, why was it not detected previously? We have better spatial resolution, which facilitates the detection, but in addition our observations were made when the quasar was significantly fainter (by 0.6 mag, according to ref. 4). From the fraction of galaxy light that we observe in the A component ($f_{G4}/f_A = 0.46$, Table 2), we calculate that the corresponding fraction was only 0.26 in 1980 and 1981. Fourth, why are the colours of A not redder than those of B or C? The best constraint is from ref. 2,

which gives the $g-r$ of A as 0.05 ± 0.06 mag bluer than that of C. An Scd galaxy at a redshift of 0.3 (see below for this choice of redshift), contributing 26% of the total light of A (the measurements of ref. 2 were made in 1980) would make A 0.09 mag redder than C in $g-r$. Similar contributions would arise from earlier type galaxies at smaller redshifts. We regard these colour data as the most stringent requirement that the putative galaxy must meet, but given the dispersion of ~ 0.1 mag in the colours of a galaxy as a function of its Hubble type, such a galaxy is not ruled out by existing data.

In addition to the fitting procedures described above, we have attempted to detect the imaging galaxy directly, by removing the effects of the seeing in the data using a nonlinear deconvolution procedure. The deconvolution procedure used was the Jansson-Van Cittert algorithm as described by Heasley⁹. The results of the deconvolution are shown in Fig. 3. Although A₁ and A₂ are resolved from each other for the first time, there is only a hint of structure aligned with the major axis of G4 (compare Fig. 1c). The emission between components A and B, visible in the raw data (Fig. 1a), remains after the deconvolution, possibly indicating yet another object.

We made spectroscopic observations of galaxies G1 and G2 (ref. 2) on 25 March 1985 UT with the Hawaii grism (grating/prism) spectrograph and Galileo/IFA CCD (charge-coupled device) mounted at the Cassegrain focus of the Canada-

Table 3 Rest-frame properties of G4 and M104

Galaxy	$\mu_v(0)$ (mag (arc s) ⁻²)	Core radius (kpc)	M_v	ϵ	Type
M104*	15.3	1.1†	-23.2	0.3	Sa
G4 ($z = 0.3$, $q_0 = 0$)	15.2‡	1.4/h§	-23.3+ 5 log ₁₀ h‡§	0.4	?

M_v , absolute visual magnitude; q_0 , deceleration parameter.

* All properties from ref. 13.

† Derived from a de-Vaucouleurs-law effective radius of 14.2 kpc using core radius = effective radius/13.2 (ref. 14).

‡ Assuming the brightness of component C was unchanged from ref. 4 (our crude photometry using the sky as a zero point shows this assumption to be true to within 0.2 mag) and using a k -correction appropriate to an Sab galaxy of 0.76 mag.

§ Hubble constant $H_0 = 100 h$.

|| Core light only.

France-Hawaii Telescope. Exposures of 57 and 31 min were made of G1 and G2, respectively, through holes in thick clouds. The spectra are normal. Reshifts derived from spectral lines Ca II H, K, H δ and G for G1 and the 4,000-Å break for G2 are 0.306 ± 0.002 and 0.304 ± 0.005 , respectively. Thus these galaxies form a small group at redshift 0.3, which may also contain G3.

It is possible that G4 is the brightest member of this group. Table 3 lists the properties of G4 if it is at redshift $z = 0.3$ and compares those properties with those of M104, a massive spiral galaxy. From the simulations presented by Schweizer¹⁰, the seeing corrections to the core radius and central surface brightness are $\sim 20\%$ if G4 follows a King model and the point spread function is gaussian. We have not applied these corrections to our data because our point spread function is not gaussian, and corrections of this order will not affect our conclusions. G4 and M104 are very similar. Of course, the redshift of G4 may not be the same as that of G1 and G2, but until a spectrum can be obtained, this is a reasonable assumption.

It is obvious that G4 will contribute substantially to the gravitational focusing. If it is the brightest member of the group, it is a massive galaxy. However, the rest of the group must also make some contribution. The only quantitative models¹¹ of the gravitational focusing explain PG1115+08 using a single flattened galaxy roughly equidistant from all four components, rather than midway between A₁ and A₂. However those models did not incorporate the galaxy group, so it is as yet unclear whether their combined effects can produce the observed configuration. A spectrum of G4 and additional lens modelling, particularly incorporating a possibly nonspherical group, are required for further progress.

After this paper was written we received a preprint from Shaklan and Hege¹² in which they report a flux excess above

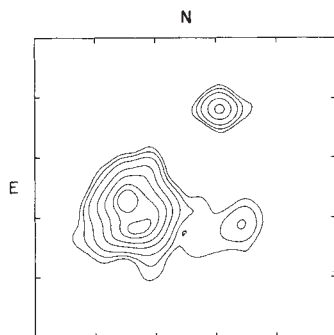


Fig. 3 Contour diagram of the PG1115+08 field after image deconvolution. The tick marks along the axes are at 1 arc s intervals. Contours are at 0.9, 0.8, 0.6, 0.4, 0.2, 0.1 and 0.05 of the peak.

that from the four usual components that is located at $\sim 65\%$ of the distance between A and B and closer to B. We can apparently confirm this report, (see Fig. 3), although our nominal brightness is greater by at least 1 mag. Our resolution of the component may explain the small value of f_B/f_A that we observe.

This research was supported in part by University of Hawaii Research and Training Fund grants to Alan Stockton to adapt the rapid-exposure film camera for use with the image intensifier, and to J.P.H. for observing expenses. We thank John MacKenty for writing the PDS scanning programs, and Robert Hlivak for keeping our CCD in good condition. J.P.H. was a Visiting Astronomer at the Canada-France-Hawaii Telescope, operated by Canadian NRC, the French CNRS and the University of Hawaii.

Received 9 December 1985; accepted 25 February 1986.

1. Weyman, R. J. *et al.* *Nature* **285**, 641-643 (1980).

2. Young, P. *et al.* *Astrophys. J.* **244**, 723-735 (1981).

3. Hege, E. K., Angel, R. P., Weyman, R. J. & Hubbard, E. N. *Nature* **287**, 416-417 (1980).

4. Vandieriet, C. *et al.* in *Quasars and Gravitational Lenses*, Proc. 24th Liege astrophys. Colloq. (1983).
5. Hege, E. K., Hubbard, E. N., Strittmatter, P. A. & Worden, S. P. *Astrophys. J. Lett.* **248**, L1-L3 (1981).
6. Foy, R., Bonneau, D. & Blazit, A. *Astr. Astrophys.* **149**, L13-L16 (1985).
7. King, I. R. *Publ. astr. Soc. Pacif.* **83**, 199-201 (1971).
8. King, I. R. *Astr. J.* **71**, 64-75 (1966).
9. Heasley, J. N. *Publ. astr. Soc. Pacif.* **96**, 767-772 (1985).
10. Schweizer, F. *Astrophys. J.* **233**, 23-34 (1979).
11. Narasimha, D., Subramanian, K. & Chitre, S. M. *Mon. Not. R. astr. Soc.* **200**, 941-950 (1982).
12. Shaklan & Hege, E. K. Preprint no. 609, Steward Observatory, Tucson, Arizona.
13. Boroson, T. *Astrophys. J. Suppl. Ser.* **46**, 177-209 (1981).
14. Kormedy, J. *Astrophys. J.* **218**, 333-346 (1977).

An apparent gravitational lens with an image separation of 2.6 arc min

Edwin L. Turner*, Donald P. Schneider†, Bernard F. Burke‡, Jacqueline N. Hewitt‡, Glen I. Langston‡, James E. Gunn*, Charles R. Lawrence§ & Maarten Schmidt§

* Princeton University Observatory, Princeton, New Jersey 08544, USA

† Institute for Advanced Study, Princeton, New Jersey 08540, USA

‡ Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

§ Department of Astronomy, California Institute of Technology, Pasadena, California 91125, USA

We report here observations which confirm Paczynski's speculation¹ that the previously known² pair of 19th magnitude quasars 1146+111B,C are actually two images of a single object produced by a gravitational lens. The image splitting is 157 arc s, more than 20 times greater than any previously reported³, thus indicating an exceptionally massive lensing object. The data supporting the lens hypothesis are remarkably similar, high signal-to-noise, moderate-resolution spectra of the two components. Both spectra show strong Mg II $\lambda 2,798$ emission at $z = 1.012 \pm 0.001$ with indistinguishable redshifts ($\Delta v = 126 \pm 309$ km s⁻¹), widths (FWHM = 64 ± 4 Å) and detailed profile shapes. Both spectra also show broad troughs at $\lambda 6,180$ and several weaker continuum features. Neither object exhibits the [O II] $\lambda 3,727$ line which is frequently strong in Mg II emission objects. If the foreground galaxy clustering apparent in a deep R band charge-coupled device (CCD) image proves insufficient to explain the large image splitting, other possibilities such as massive dark objects (for example, a $\sim 10^{15} M_{\odot}$ black hole) or a cosmic string may be indicated.

In 1979, Hazard *et al.*⁴ reported a 'concentration' of quasars in a field at $\alpha = 11$ h 46 min 14 s and $\delta = 11^{\circ}11'42''$ (1950). Subsequently, Arp and Hazard² gave redshifts based on slit spectra for nine confirmed quasars in the field. Among these, two quasars (B and C) were determined to have redshifts of 1.01, equal within their measuring errors. Arp and Hazard² dismissed the possibility that these were a pair of gravitational lens images on the grounds of an implausibly large image separation. Recently, Paczynski¹ again called attention to this system along with four others as possible examples of large splitting gravitational lenses and noted that such systems might result from an intervening cosmic string^{5,6}.

We obtained spectra of 1146+111B and C on the night of 5 March 1986 UT with the Cryogenic Camera⁷ and the R-C spectrograph on the Kitt Peak National Observatory 4-m telescope of the National Optical Astronomy Observatories (NOAO). A 2.5 arc s $\times$ 4.4 arc min slit was used to illuminate a 300-line per millimetre grism (grating/prism). An 800 $\times$ 800 TI CCD camera recorded the first-order spectra through a GG420 filter giving 15 Å resolution in the $\lambda \lambda 4,600$ -7,800 range with a 4.3-Å per pixel sampling. Two 1,800-s exposures of each object

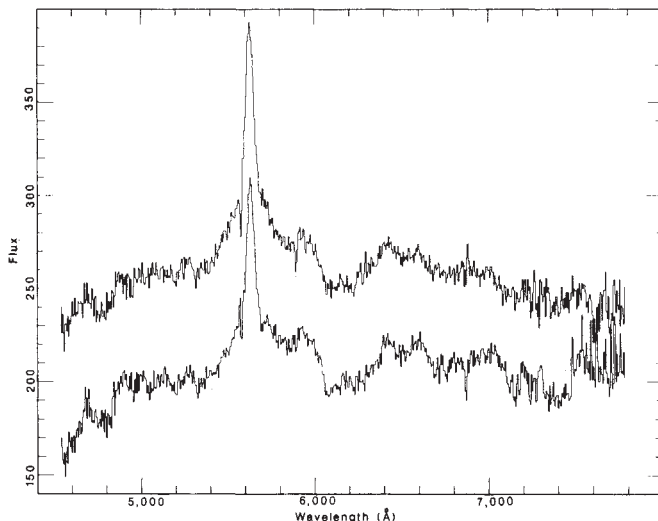


Fig. 1 The spectra of 1146+111B and C. The vertical axis is a linear flux scale in arbitrary units. The upper spectrum is of object B; the lower, of object C (70 counts per channel were added to B to displace it from C). The identification of the strong emission feature as Mg II λ 2,798 is primarily based on Arp and Hazard's² detection of C III λ 1,909. Note the correspondence of the emission feature, the deep trough, and several weaker continuum dips and bumps in the two spectra.

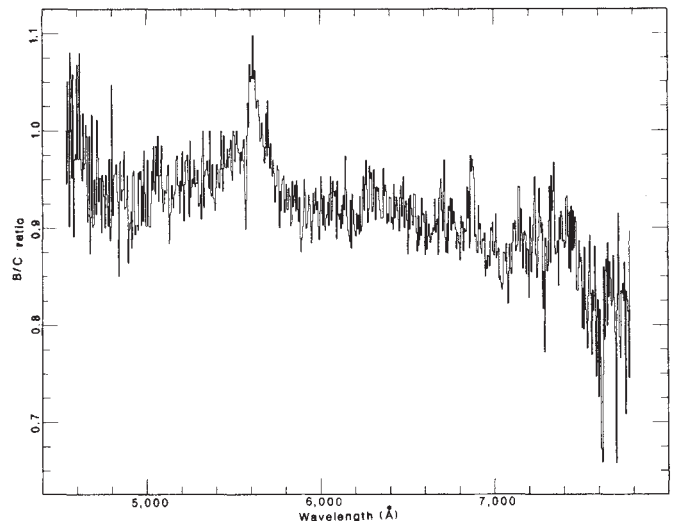


Fig. 2 The ratio of the spectrum of 1146+111B to that of 1146+111C. The feature in this ratio at the wavelength of the Mg II emission indicates that the line-to-continuum ratio in the two objects is slightly different. The smooth variation in the ratio from blue to red indicates that C has a redder detected continuum than B. The absence of other significant features in this ratio demonstrates the detailed match in shape and amplitude between respective emission and absorption features in the two spectra.

were obtained along with interspersed quartz lamp flat fields and He-Ne-Ar comparison arcs. A spectrum of the flux calibrator GD 140 was also obtained to remove the instrumental response function and atmospheric absorption bands. The data frames were reduced individually using the CASSANDRA image processing package and the resulting spectra of each object were then combined. They are shown in Fig. 1.

These two spectra agree precisely in their significant features. Figure 2 gives the ratio B/C of the two spectra. The Mg II line ratio is ~ 1.02 while the continuum ratio varies from roughly 0.96 at the blue end to around 0.83 at the red end. These small variations do not mitigate against a lens interpretation as similar small differences in line strengths and continuum slopes are observed in such confirmed lens cases as 0957+561 (ref. 8) and 2016+112 (ref. 9). They may be accounted for by a variety of explanations including source variability plus differential time delays, gravitational mini-lensing events¹⁰, extinction by intervening material, contamination of the source spectrum by light from the lens, and systematic observational errors resulting from differential refraction effects. Overall, the spectroscopic similarity of 1146+111B and C is as strong as for any of the known (few arc second separation) lens systems.

On 7 March 1986 UT we obtained six direct images of the 1146+111B, C field, each exposed for 300 s using the TI2 CCD camera at the prime focus of the same telescope. The average of these six images is shown in Fig. 3. Profile fits to stellar images indicate a FWHM seeing of 1.2 arc s. Although accurate calibration of this image was not attempted due to non-photometric observing conditions, we believe that it reaches 23rd to 24th magnitude in the R band.

Analysis of this image shows: (1) an image separation of 156.8 ± 0.5 arc s, (2) an R-band B/C brightness ratio of 0.93 ± 0.01 ; (3) a possible very distant rich cluster of galaxies between the two quasar images; and (4) an unusual number of fairly bright ($R \sim 18$ to 19), apparently low redshift ($z \leq 0.2$) galaxies in the field including one roughly between B and C. The quasar images themselves show no signs of resolution.

The importance of 1146+111B, C lies in its unprecedented large image separation and the questions and possibilities thus raised.

The first question is that of the nature of the lensing object. The most conventional candidate would be a very massive rich cluster of galaxies. If such a cluster lay at the most probable redshift of 0.32 and had an isothermal sphere mass distribution, the required one-dimensional velocity dispersion would lie near $2,200 \text{ km s}^{-1}$ (depending weakly on q_0). If the cluster instead lay at a redshift of 0.81, which would minimize (for a constant M/L value) its visibility (assuming an elliptical galaxy k correction¹¹), the requirement would be $\sim 4,500 \text{ km s}^{-1}$. The former value is extreme but not entirely beyond the observed range for rich clusters¹². Nevertheless, such a cluster would be easily detectable in deep CCD images at any redshift less than 1, so this candidate lens is subject to direct observational test. If no suitable rich cluster can be located, more exotic possibilities such as $\sim 10^{15} M_{\odot}$ black holes and cosmic strings will need to be considered. Indeed, lens systems much like 1146+111B, C are a predicted consequence of cosmic strings^{1,5,6}. Of course, the possibility of a physically distinct (not lensed) pair of objects with astonishingly similar spectroscopic properties is not logically excluded by the data.

Second is the possibility that additional objects in the 1146+111 field may be multiply imaged. The large angular splitting of the B, C pair implies a large multiple imaging cross-section for most reasonable lens scenarios. Some of the other known² quasars in the field could well have additional, so far undetected images; D, E and K are particularly likely candidates. Even if they are not multiply imaged, lens magnification and hence flux enhancement could account for the statistically unlikely 'concentration' of quasars which originally called attention to the field. In any case, discovery of additional multiply-imaged objects would offer the opportunity to model the lens mass distribution more precisely than for other known lenses and, should this distribution prove simple, to carry out powerful cosmological tests.

Third is the large differential time delay expected for large splitting image pairs. For reasonable lens models, the 1146+111B, C time delay is of the order of 10^3 years. This offers the opportunity of observing variations in the lensed object(s) taking place over time periods well in excess of those usually accessible. If the lensing geometry is highly symmetrical, as the roughly

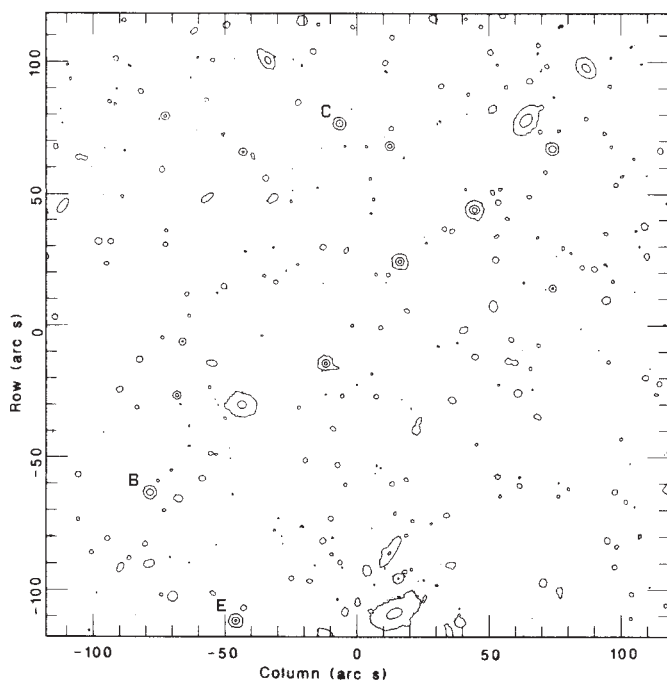


Fig. 3 A contour diagram of the R-band CCD frame of the 1146+111 field. North is at the top, east at the left. The image is 4 arc min on a side. The first contour level is 2.5 times the r.m.s. sky noise with additional contours at 3-magnitude intervals. Dots corresponding to single high pixels are probably noise in many cases; however, larger objects are likely to be real. It is estimated that the faintest objects visible are between 23rd and 24th *R* magnitude. Object B is at column, row coordinates -77.4 arc s, -61.9 arc s while object C is at -6.0 arc s, 77.7 arc s. Note that there are several bright galaxies scattered across the frame including one at -42.5 arc s, -28.7 arc s (roughly between B and C) and a concentration near the centre of the south edge of the field. Also note the arc of faint galaxies, possibly a distant cluster, extending from about -5 arc s, 26 arc s towards -74 arc s, -10 arc s.

equal brightnesses of B and C might suggest, the time delay will be substantially smaller than this 'natural' value.

1146+111B, C and the surrounding field warrant further study, and some attempts in this direction are under way. Verification of the spectral similarity of B and C at other wavelengths and searches for additional, multiply-imaged objects are of particular interest. Other candidate large splitting lenses should also be examined more carefully.

We thank B. Paczynski for a preprint and for discussions, the Kitt Peak staff for assistance with the observations and J. Ostriker for useful comments on the manuscript. D.P.S. acknowledges the support of an Exxon Fellowship. This work was supported by NSF grants AST84-20352 and AST83-14134 and NASA grants NAGW-765 and NAS5-29225. The NOAO is operated by the Association of Universities for Research in Astronomy, Inc., under contract with the NSF.

Received 24 March; accepted 10 April 1986.

1. Paczynski, B. *Nature* **319**, 567-568 (1986).
2. Arp, H. & Hazard, C. *Astrophys. J.* **240**, 726-736 (1980).
3. Turner, E. L. *IAU Symp.* No. 117, 209-218 (1986).
4. Hazard, C., Arp, H. & Morton, D. C. *Nature* **282**, 271-272 (1979).
5. Vilenkin, A. *Astrophys. J. Lett.* **282**, L51-L53 (1984).
6. Gott, J. R. *Astrophys. J.* **288**, 422-427 (1985).
7. DeVney, J. *An Observer's Manual for the Cryogenic Camera* (Kitt Peak National Observatory, Tucson, 1985).
8. Young, P., Gunn, J. E., Kristian, J., Oke, J. B. & Westphal, J. A. *Astrophys. J.* **241**, 507-520 (1980).
9. Lawrence, C. R. *et al. Science* **223**, 46-49 (1984).
10. Paczynski, B. *Astrophys. J.* **301**, 503-516 (1986).
11. Schneider, D. P., Gunn, J. E. & Hoessel, J. G. *Astrophys. J.* **264**, 337-355 (1983).
12. Mushotzky, R. F., Serlemitsos, P. J., Smith, B. W., Boldt, E. A. & Holt, S. S. *Astrophys. J.* **225**, 21-39 (1978).

Structures in the coma of comet Halley

E. Grün*, U. Graser†, L. Kohoutek‡, U. Thiele§,
L. Massonne*|| & G. Schwehm¶

* Max-Planck-Institut für Kernphysik, D-6900 Heidelberg 1, FRG

† Max-Planck-Institut für Astronomie,
D-6900 Heidelberg-Königstuhl, FRG

‡ Hamburger Sternwarte, Gojenbergsweg 112,
D-2050 Hamburg 80, FRG

§ German-Spanish Astronomical Center, Calar Alto,
Apartado Correos 511, 04080 Almeria, Spain

|| ESOC/ECD, Robert-Bosch-Str. 5, D-6100 Darmstadt, FRG

¶ ESTEC/SSD, Postbox 299, 2200 AG Noordwijk, The Netherlands

Dense dust jets are expected to be encountered by the spacecraft penetrating the coma of comet Halley in March 1986. At the time of encounters, remote observations of the global distribution of coma constituents will complement the detailed *in situ* measurements. Here we report ground-based measurements of the optical brightness distribution of the coma, made in November 1985, when the comet was at a distance of 1.75 AU from the Sun. From the total brightness we deduce a dust production rate of 1.5×10^6 g s⁻¹, but our results show that cometary dust and gas production, which seem to be correlated, are strongly anisotropic and time-variable. Structures in the coma can be revealed by comparison with a physical model of the coma, but astrometry of the comet nucleus is adversely affected by the anisotropic brightness distribution in the coma.

The coma morphology and dust emission pattern of comet Halley, as revealed on photographic plates taken during the 1910 apparition¹, have been used to determine the nucleus spin vector². A verification of this result and the location of active regions on the nucleus is important for the space probe missions Giotto³ and Vega⁴ currently approaching the comet. An efficient computational method has been developed⁵ to analyse images of the coma and to model the dust distribution in it. Repeated observations of coma structures are necessary to determine the rotational state of the nucleus and the distribution of active regions on it.

The observations were performed using the 80-cm Schmidt telescope and the 2.2-m telescope at the German-Spanish Astronomical Center at Calar Alto, Spain. Sixteen Schmidt plates were taken during the period 9-21 November 1985, in order to study the brightness distribution of the outer coma. The following emulsion-filter combinations were used: Kodak IIIa-J plus Schott UG1 or GG495, IIIa-F plus RG610, or 103a-E without filter.

A large, curved, jet-like structure extending 21 arc min to the south of the cometary nucleus was photographed on two plates on the night of 12/13 November; the U-plate image is shown in Fig. 1. The observed structure was nearly perpendicular to the projected antisolar direction; although the projection angle is unknown, this deviation seems too large to be only a projection effect. The brighter inner part of the jet (visible on Fig. 1) is narrow and consists of three 'beams'; its wide and faint outer part looks like a fan. The U-plate (IIIa-J + UG1) has shown that this ejection very probably contained not only dust particles but also CN molecules. Jockers⁶ observed a southward-pointing ion tail (CO⁺) during the same night: it is thus highly likely that ion emissions contribute to the total brightness observed by us. The coma photographed 3 days earlier was symmetrical, and had a diameter of ~ 10 arc min (at the isophote level corresponding to the sky background). In addition to the jet, some radial rays are visible on the U-plate at the west side of the coma. On 18.0 November the outer coma appeared symmetrical again, but had expanded to ~ 15 arc min. In addition to the increase of the cometary brightness due to its decreasing distance from both the Earth and the Sun, the activity observed on 12/13 November

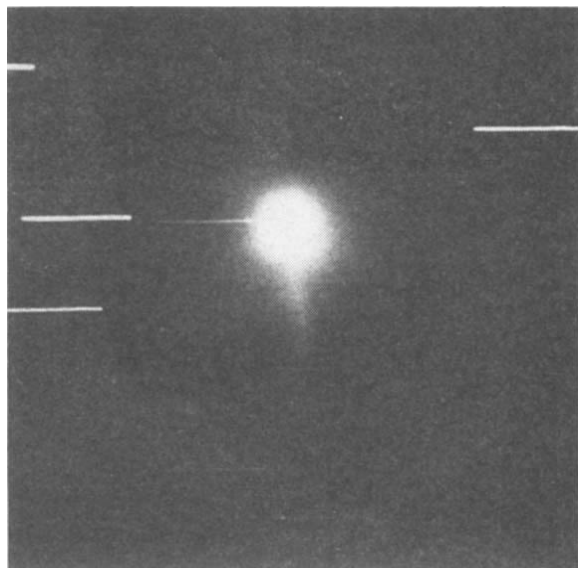


Fig. 1 Jet structure of the coma of comet Halley, observed on 13.0 November 1985 UT (Universal Time) with the 80-cm Schmidt telescope at Calar Alto. The 50-min exposure was taken on Kodak IIIa-J emulsion through a UG1 filter. The field of view is 20×20 arc min; north is up, east is left.

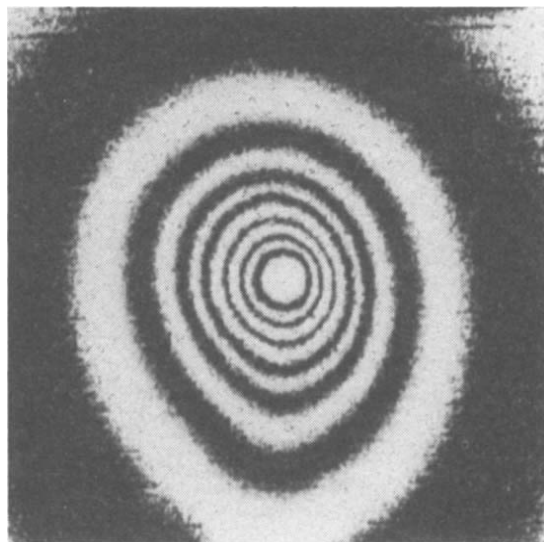


Fig. 2 Isophote representation of a CCD image taken with the 2.2-m telescope at Calar Alto on 12.9 November 1985 UT. A continuum filter at 703 nm was used. The field of view is 2×2 arc min; north is up, east is left.

probably increased the total brightness of the comet during the subsequent days. On plates taken on 21.0 November two jets are visible to the east (position angles 63° and 90° ; the latter is the tail direction), at a distance of ~ 9 arc min from the nucleus.

More than 100 CCD (charge-coupled device) images have been taken with the 2.2-m telescope. The detector used was a thinned, backside-illuminated RCA CCD (type SID 510-EX), with 512×320 pixels. The pixel size is $30 \times 30 \mu\text{m}$, which corresponds to 0.35 arc s at the Cassegrain focus. The filters used were Johnson R , V and B and the IHW filters at 703 nm ($\Delta\lambda = 22$ nm), 684 nm ($\Delta\lambda = 9$ nm), 514 nm ($\Delta\lambda = 10$ nm) and 485 nm ($\Delta\lambda = 6$ nm). The filter at 514 nm covers a major C_2 emission; the other narrow ones are continuum filters.

Optimum pre-perihelion observation conditions (small zenith distance, large brightness and long observation intervals from a single site) occurred in November 1985. During this time the comet was ~ 0.7 AU from the Earth, in approximately the anti-solar direction. The comet was close to its ascending node. The required high offset speed of up to 400 arc s h^{-1} was achieved with an accuracy of better than 3 arc s h^{-1} . The exposure times to reach near-saturation in the centre ranged from 15 to 1,000 s for the wide-band and narrow-band filters, respectively.

Figure 2 shows an isophote representation of a CCD image of the coma. The image was taken on 12.9 November. The outer part of the image is elongated towards the south, whereas the centre part is more symmetrical. An image taken ~ 3.5 h after that shown in Fig. 2 shows that the direction of strongest emission had shifted by 10° to position angle 190° . A similar shift was observed for the jet seen on the Schmidt plates which were taken at about the same times.

To determine photometric magnitudes for Halley's central coma, a set of standard star measurements were taken in the same (photometric) night (~ 12 stars per waveband with magnitudes between 8.8 and 17. The sky background (per pixel and second) in the Halley frames was determined by averaging the background of three field exposures taken at about the same zenith distance, shortly before or after the respective Halley exposure. Coma magnitudes were measured by flux integration in circles of 10 and 30 arc s diameter around the brightness maximum. The measured values are $B = 12.6$ and 11.3 ,

$V = 11.6$ and 10.4 and $R = 11.4$ and 10.2 , respectively. A conservative estimate of the root-mean-square deviation of one observation is ± 0.1 mag. The magnitudes were measured on 13.1 November.

Assuming that most of the coma brightness in R can be attributed to sunlight scattered by dust, one can derive the dust production rate. For the measured brightness in R of 11.4 mag we find a mean dust production rate of $\sim 1.5 \times 10^6 \text{ g s}^{-1}$. We adopt an effective geometric albedo of 0.03, a grain size distribution with an exponent -4.2 in the coma and the effective emission velocity of 150 m s^{-1} according to the standard Halley model parameters proposed by Newburn and Reinhard⁷. Note, however, that the uncertainties in the assumed parameters result in an uncertainty of a factor of ~ 2 in the production rate.

Taking an average dust-to-gas ratio of 0.3 and a mean molecular mass of 22.3 AMU (ref. 7) we derive a gas production rate of $\sim 1.3 \times 10^{29} \text{ molecules s}^{-1}$. This is ~ 3 times higher than the value predicted by Newburn⁸ on the basis of 1910 Halley observations and recent cometary models.

Figure 3 shows radial and azimuthal brightness profiles of the Halley coma. The radial brightness profiles at different wavelengths were obtained by averaging the surface brightness around concentric circles centred on the brightness maximum. Inside $\rho \approx 5$ arc s (ρ being the distance from the centre) the brightness first rises more steeply than the average increase, then flattens off at $\rho < 3$ arc s. This inner part of the coma represents the seeing disk of the central brightness peak. During the observation period the seeing ranged from 2 to 5 arc s. Outside this radius the coma brightness in the continuum wavelengths decreases roughly as $1/\rho$, as would be expected from a steady stream of particles emitted isotropically at constant speed from the surface of the nucleus. In the outer part of the brightness profile, the effect of the constant sky background is recognized. The radial profiles of the C_2 brightness fall off as $\rho^{-\gamma}$, $\gamma = 0.6\text{--}0.75$. This is in good agreement with the C_2 scale length calculated for the inner coma region⁹.

The azimuthal brightness profile at continuum wavelengths (Fig. 3b) shows deviations of up to 30% from the mean brightness at a given distance from the centre. The C_2 brightness varies azimuthally in the same way, with a smaller amplitude.

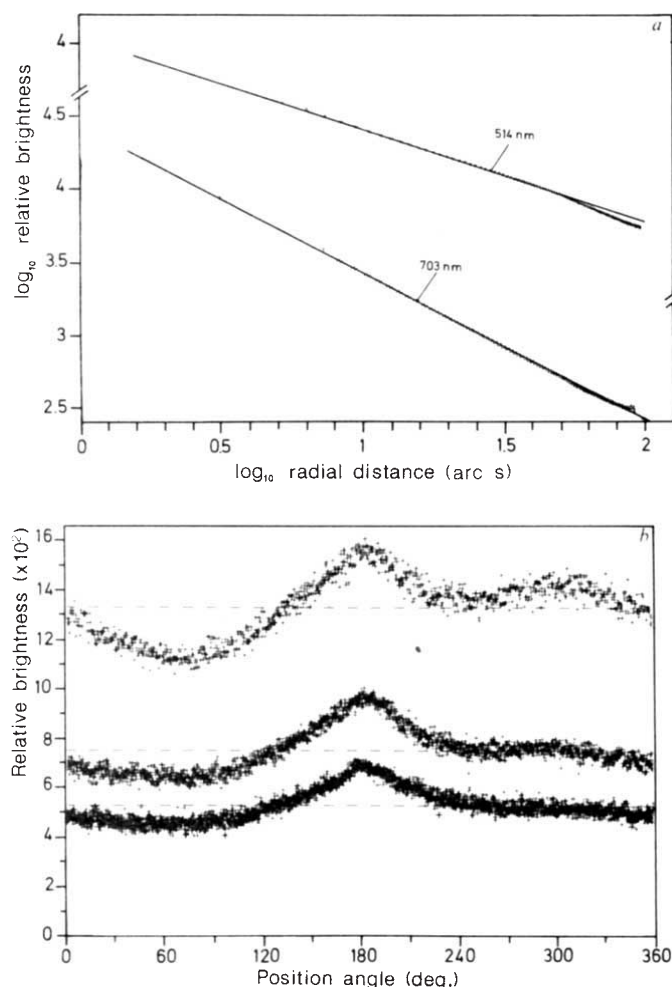


Fig. 3 Brightness profiles of CCD images of the Halley coma, taken on 12.9 November 1985 UT. *a*, Radial profiles of the continuum brightness (703 nm) and C₂ brightness (514 nm). The solid lines have slopes of -1.0 and -0.63 , respectively. *b*, Azimuthal profiles of the brightness at 703 nm at radial distances of 21, 37 and 53 arc s from the brightness centre. The dashed lines represent mean brightness values at these distances.

The strong overall radial brightness gradient hinders the recognition of local variations in brightness. To enhance the visibility of local inhomogeneities, and hence the variations in dust density, we have compared the images with a simple physical model of the dust coma. After subtraction of a constant background level the images were normalized to a $1/\rho$ brightness distribution centred on the apparent nucleus. Thus, the seeing disk is represented by a low-intensity central part and a bright ring around it. Outside this artefact the brightness distribution represents the deviation from an isotropic expanding dust emission. These deviations are due to non-isotropic emissions and bending of trajectories caused by dynamical effects. A similar method can be used to show structures in the gas coma.

Figure 4 shows a processed image, taken on 21.1 November. It also shows the main dust emission from the nucleus in the southerly direction. This time, however, it bends over to the east, which is the general tail direction. The enhancement of the outer coma in the easterly direction became increasingly dominant during the latter part of the observation period. At position angle 63° , a narrow jet is visible which continues for several arc minutes and has been simultaneously observed on Schmidt plates as described above.

We have found several independent indications that enhancements in the inner dust coma are correlated with enhancements

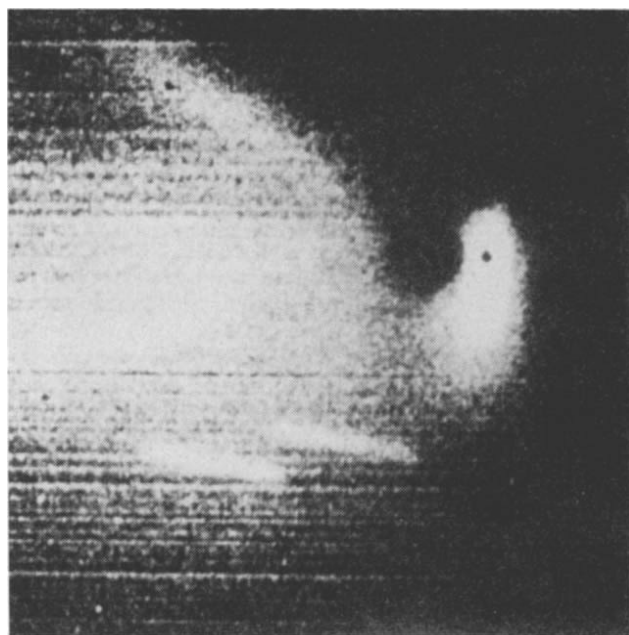


Fig. 4 Jets in Halley's inner dust coma on 21.1 November 1985 UT. This picture (at 703 nm) was obtained by normalizing the observed brightness distribution to the distribution which would be expected for an isotropically expanding coma. The centre of the original brightness distribution is marked by a dark spot in the right-hand part of the image. A major jet can be seen to the south, which bends over to the east. Another jet is visible at position angle 63° . Two star trails can be recognized in the lower part of the image. The east-west striation originates from a defect in the CCD, which could not be corrected for in all portions of the picture. The field of view is 2×2 arc min; north is up, east is left.

in the gas coma, as well as with jets and perhaps even ion tails outside the coma. This observation indicates that the interaction between dust and gas, and perhaps ions, is more complicated than previously expected.

On nearly all images taken with continuum filters, a prominent emission feature is visible within $\sim 10^\circ$ of south. This feature shows no periodic orientation changes, such as are generally expected from an active area on a rotating nucleus. We suggest, therefore, that it is caused by increased gas and dust emission at the constantly sunlit rotational pole of the nucleus. Each observation then defines a plane in which the spin vector must be located. All of these planes should intersect in a common line, which is the spin vector. For three representative observations on 12/13 November and three others on 20/21 November these planes have been computed. Their common spin vector lies in the region $\phi = 0-10^\circ$ and $I = 30-55^\circ$, following the notation of Sekanina and Larson². This differs by $\sim 40^\circ$ from the direction recently proposed by Sekanina¹⁰ on the basis of 1910 observations.

The anisotropic brightness distribution of the coma affects the determination of the nucleus position. On some images we have observed elongated seeing disks, suggesting that the deviation of the brightness centre from the nucleus position may be as large as one-third of the mean seeing-disk radius. Local dust density enhancements by about two orders of magnitude—as required to produce visible structures, according to jet simulations⁷—are also expected to occur during the space probe encounters with the comet. The observations cannot tell us whether hazardous large particles (> 1 mg) are entrained in the jets, because the observed brightness is mainly due to 1–10- μ m-sized particles for all reasonable size distributions.

Remote ground-based observations of the coma, similar to

those presented here, will be of great value during the spacecraft encounters with comet Halley. The remote coma observations will determine the overall structure, while *in situ* measurements provide the absolute calibration.

We thank T. A. Morley, ESOC, for providing exact Halley ephemeris data. E.G. and L.K. are visiting astronomers, German-Spanish Astronomical Center, Calar Alto, operated by the Max-Planck-Institut für Astronomie jointly with the Spanish National Commission for Astronomy.

Received 13 January; accepted 10 March 1986.

1. Larson, S. M. & Sekanina, Z. *Astr. J.* **89**, 571-578 (1984).
2. Sekanina, Z. & Larson, S. M. *Astr. J.* **89**, 1408-1425 (1984).
3. Reinhard, R. *Adv. Space Res.* **2**, 97-107 (1983).
4. Sagdeev, R. Z. *et al. Adv. Space Res.* **2**, 83-90 (1983).
5. Massonne, L., Fertig, J., Grün, E. & Schwehm, G. in *Asteroids, Comets, Meteors II* (eds Lagerkvist, C.-I., Lindblad, B. A., Lundstedt, H. & Rickman, H.), 407-410 (in the press).
6. Jockers, K. & Geyer, E. H. *The ESO Messenger* No. 42, 9 (1985).
7. Newburn, R. L. & Reinhard, R. *Eur. Space Ag. spec. Publ.* SP-174, 19-24 (1981).
8. Newburn, R. L. *ESA Eur. Space Ag. Spec. Publ.* SP-174, 3-18 (1981).
9. Cochran, A. L. *Astr. J.* **90**, 2609-2614 (1985).
10. Sekanina, Z. *IAU Circ.* No. 4151 (1985).

Nitrogen dioxide emitted from space shuttle surfaces and shuttle glow

Ulf von Zahn* & Edmond Murad†

* Physikalisches Institut, Universität Bonn, 5300 Bonn 1, FRG

† Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts 01731, USA

Photographs taken on early space shuttle flights have indicated the presence of a surface glow in the ram direction¹. Examination of airglow data from the AE and DE satellites has shown the presence of a similar glow², suggesting that the phenomenon is general to spacecraft in low Earth orbit, with the emission being concentrated in the ram direction. Several explanations have been offered for the glow: (1) OH formed by surface reactions of fast O atoms³; (2) plasma discharge mechanism⁴; (3) surface-induced decomposition of N₂ followed by recombination into an excited state⁵; (4) surface emission by impact of O₂ and O (ref. 6); and (5) near-continuum emission (at a resolution of ~3.4 nm) in the region 400-800 nm from electronically excited NO₂ (ref. 7). With reference to the last hypothesis we discuss here data on the number density of NO₂ measured inside the payload bay of the shuttle orbiter. We also derive fluxes of NO₂ near shuttle surfaces. Our data are shown to be consistent with the hypothesis of the glow being caused by NO₂ emission.

Number densities of NO₂, [NO₂], were measured with a neutral mass spectrometer aboard flight 41-B of the space shuttle. The instrument and its performance have been described in detail by Wulf and von Zahn⁸. The instrument was mounted near the middle of the payload bay. The data discussed here were obtained at the beginning of day 6 of the mission, when the ion source opening was pointing towards the pressure bulkhead of the astronaut cabin. Figure 1 shows the motion of the orbiter and the position of the mass spectrometer sampling orifice. Incoming ambient species bombarded the cabin bulkhead directly; however, the mass spectrometer observed only species reflected or desorbed from the shuttle surfaces, mostly from the pressure bulkhead. The orbiter maintained the attitude shown in Fig. 1 for 11 min, enough time for the signals measured in the mass spectrometer to reach a steady state. The following data were obtained at the end of this period at 290 km altitude and 27° latitude, near 03:45 LT (Local Time). [NO₂] is derived

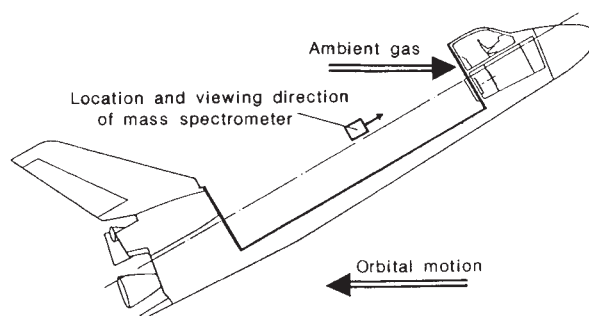


Fig. 1 Observational geometry for measurement of NO₂ density in the payload bay of the shuttle orbiter. The ambient gas strikes the orbiter surfaces with a velocity of 7.3 km s⁻¹, but the ion source of the mass spectrometer receives only species reflected or desorbed from orbiter surfaces.

from the peak at 46 AMU in the mass spectrum, after correcting for the contribution by ¹²C¹⁶O¹⁸O (heavy CO₂), deduced from the peak at 44 AMU. We find that 43% of the observed 46 AMU peak is due to heavy CO₂. The remaining 57% we assign to NO₂ (the contribution of monomethylhydrazine to the mass 46 peak is considered to be negligible, mainly because of the strong dependence of the observed mass 46 peak on the angle of attack). To derive an absolute value for [NO₂], we use the absolute sensitivity of the mass spectrometer for CO₂, as calibrated in the laboratory, multiplied by the ratio of cross-sections for electron impact ionization of NO₂ and CO₂ (refs 9, 10). This yields the NO₂ density inside the ion source of the mass spectrometer, [NO₂]_s = (2.7 ± 1.0) × 10⁷ cm⁻³. With an ion source temperature of 315 K, the flux Φ(NO₂)_o out of the ion source opening becomes (2.5 ± 1) × 10¹¹ cm⁻² s⁻¹. The error bars reflect our estimate of an upper bound for potential contributions to the mass 46 peak from gases other than NO₂ and CO₂.

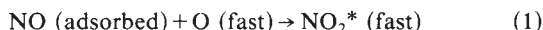
In the presumed steady-state conditions, the flux into the ion source, Φ(NO₂)_i, equals the flux out of it, Φ(NO₂)_o. Because of the observation geometry, Φ(NO₂)_i is only ~40% of Φ(NO₂)_b, the flux leaving the cabin bulkhead; hence, Φ(NO₂)_b ≈ (6 ± 3) × 10¹¹ cm⁻² s⁻¹.

To relate this flux to a density [NO₂]_b near the bulkhead we consider two extremes which have been suggested by other studies of the shuttle glow. First, NO₂ desorbs from the bulkhead with the mean thermal velocity (0.37 km s⁻¹). This results in [NO₂]_b = 6 × 10⁷ cm⁻³. Second, NO₂ desorbs with a velocity of 2.5 km s⁻¹ (corresponding to a kinetic energy of 1.5 eV), as suggested by Swenson *et al.*⁷, based on a calculated *e*-folding distance of 20 cm (ref. 11) for the shuttle glow and a radiative lifetime of NO₂^{*} of 80 μs (ref. 12). This velocity would lead to a considerable thermal accommodation factor for [NO₂] inside the ion source, and for these circumstances [NO₂]_b ≈ 1 × 10⁷ cm⁻³.

We now consider the requirements of the shuttle glow intensities, assuming that the glow arises solely from NO₂ as suggested by Swenson *et al.*⁷. The most recent estimate of the glow intensity is 10-100 kR (1R = 10⁶ photons cm⁻² s⁻¹), viewed normal to the surface¹³. This translates into a flux of excited NO₂, Φ(NO₂^{*}), of 10¹⁰-10¹¹ molecules cm⁻² s⁻¹. The surface above which this glow is calculated consists of the shuttle tiles, which are made of silica fibres doped with a variety of materials¹⁴. On the other hand, a large part of the NO₂ observed by our mass spectrometer originates from the bulkhead surfaces, which are covered by Teflon-coated glass¹⁵. Mende and Swenson¹³ reported a study of the relative glow intensities of different materials aboard shuttle mission 41-D at ~300 km altitude. Their study indicated that the glow intensity was weakest above a polyethylene sample and brightest above a material designated Z302, overcoated with silicon. If this latter material is close to that of the shuttle tiles, one would have to

draw the conclusion that the NO_2 flux close to the shuttle tiles is considerably enhanced over that near the bulkhead. Assuming that the conditions for the two flights are the same (mass spectrometer data from flight 41-B and glow data from flight 41-D), and that the glow intensity of the shuttle tiles is the same as that of Si-coated Z302, we deduce that $\Phi(\text{NO}_2^*)/\Phi(\text{NO}_2) \approx 0.01\text{--}0.1$. Thus the mass spectrometer data indicate that NO_2 desorbed from the surface is present in sufficient amount to further consider NO_2 a viable candidate for causing the glow.

To pursue this analysis, let us assume that NO is present on the surfaces of the space shuttle (from the ambient, from surface reaction between N and O (ref. 16), from chemical reaction of fast O with the shuttle materials, or from contamination caused by attitude thruster firings⁸), and that O atoms striking the surface recombine with the adsorbed NO to yield NO_2^* :



This is an example of the Eley-Rideal mechanism in catalysis (see ref. 17 for a discussion of this and the Langmuir-Hinshelwood mechanism). In this mechanism, the desorbed molecule is expected to carry the evolved heat of the reaction as rotational, vibrational, electronic or kinetic energy (or a mix of all these). A summary of work relevant to the shuttle glow problem is given in ref. 18. For the case under discussion, reaction (1), the excess energy is enormous: kinetic energy of O (4.5 eV, corresponding to a velocity of 7.3 km s^{-1}) + $D_0(\text{NO-O})$ (3.1 eV), or a total of 7.6 eV. The short wavelength cutoff of the glow is $\sim 430 \text{ nm}$ (ref. 7), corresponding to 2.9 eV in the form of electronic excitation. If the desorbed NO_2^* has a kinetic energy of 1.5 eV, as has been suggested¹³, then the remainder (3.2 eV, or $\sim 40\%$ of the excess energy) must be partitioned between vibrational/rotational excitation of NO_2 and surface deposition of the excess energy. Results from beam experiments¹⁸ and theoretical calculations¹⁹ indicate that the amount of energy deposited into the surface is quite small. For example, in a theoretical study of the reaction of O with adsorbed carbon on platinum at 0 and 500 K, it was found that $>90\%$ of the energy released in the reaction appears in the product, CO. If the case of $\text{O} + \text{NO}_{\text{ads}}$ is analogous, then most of the remaining energy would appear in vibrational/rotational excitation of NO_2 ($\sim 2.8 \text{ eV}$); $\sim 5\%$ of the excess energy would be deposited in the surface ($\sim 0.4 \text{ eV}$). The implication of this argument is that the glow emitted by NO_2 may be more intense in the infrared than in the visible. The problem becomes even more serious if the velocity of the desorbed NO_2^* is less than that used above. In any case, the prediction from this mechanism is that the rate of desorption of NO_2 will increase with increasing [O], until a plateau is reached.

Another model is provided by the Langmuir-Hinshelwood mechanism, whereby fast O atoms are thermalized on the surface and then react with adsorbed NO to form NO_2 :



Again, the NO_2 or NO_2^* would be expected to contain the excess energy, which is now considerably smaller than in the Eley-Rideal mechanism, namely only $D_0^0(\text{NO-O})$, 3.1 eV. Since the short-wavelength threshold is $\sim 430 \text{ nm}$ (2.9 eV), the remainder, 0.2 eV, is kinetic energy; this corresponds to a velocity of 0.9 km s^{-1} for the desorbed NO_2 . If we again use a radiative lifetime of $80 \mu\text{s}$, then the e -folding distance becomes $\sim 7 \text{ cm}$, which is not far from the observed distance. If both NO and O are strongly adsorbed on the surface and if the adsorption energies are comparable, then the rate of reaction (2), and therefore the glow intensity, will go through a maximum as [O] is increased. If, on the other hand, both NO and O are weakly bound to the surface, then the reaction will be second-order in [O] (ref. 17). We cannot yet choose between the two mechanisms, although the Langmuir-Hinshelwood mechanism seems easier to reconcile with the observations. Finally, the energies and lifetimes used above represent most probable values. In

reality, there will be a distribution (perhaps maxwellian) of interaction energies, which will in turn, lead to a distribution of excitation energies in the products. For the qualitative nature of this discussion, however, it is sufficient to consider the most probable values.

The question of the origin of the surface NO merits some discussion. The four sources mentioned above are all possible, and all can provide the required [NO]. We note that, for example, semi-empirical models of the thermosphere²⁰ predict ambient densities [O] and [N] of $5.3 \times 10^8 \text{ cm}^{-3}$ and $6.5 \times 10^5 \text{ cm}^{-3}$, respectively, during our observational period. Hence, fluxes of ambient species onto the cabin bulkhead have been $\phi(\text{O}) = 3.4 \times 10^{14} \text{ cm}^{-2} \text{ s}^{-1}$ and $\Phi(\text{N}) = 4.1 \times 10^{11} \text{ cm}^{-2} \text{ s}^{-1}$. $\Phi(\text{N})$ seems only just able to balance $\phi(\text{NO}_2)_b$. However, due to local-time and latitudinal/seasonal variations in the global N distribution, $\phi(\text{N})$ went through a local minimum at the location and time of our observation and the orbital average of $\phi(\text{N})$ is considerably higher.

The data reported here are consistent with a glow hypothesis based on NO_2 emission. However, they do not prove this hypothesis, and a number of questions remain, concerning NO_2 as the source of the radiation.

We thank E. Wulf for support and I. Kofsky for helpful comments. This work was supported by the German Bundesministerium für Forschung und Technologie under grant 01 QV 239.

Received 29 October 1985; accepted 4 March 1986.

1. *Aviat. Week Space Tech.* **116**, 19, 14 (1982).
2. Yee, J. H. & Abreu, V. *J. Geophys. Res. Lett.* **10**, 126-129 (1983).
3. Slinger, T. G. *Geophys. Res. Lett.* **10**, 130-132 (1983).
4. Papadopoulos, K. *Radio Sci.* **19**, 571-577 (1984).
5. Green, B. D. *Geophys. Res. Lett.* **11**, 576-579 (1984).
6. Prince, R. H. *Geophys. Res. Lett.* **12**, 453-456 (1985).
7. Swenson, G. R., Mende, S. B. & Clifton, K. S. *Geophys. Res. Lett.* **12**, 97-100 (1985).
8. Wulf, E. & von Zahn, U. *J. geophys. Res.* **91**, 3270-3278 (1986).
9. Stephan, K., Helm, H., Kim, Y. B., Seykora, G., Ramler, J. & Grössl, M. *J. chem. Phys.* **73**, 303-308 (1980).
10. Märk, T. D. & Hille, E. *J. chem. Phys.* **69**, 2492-2496 (1978).
11. Yee, J. H. & Dalgarno, A. in *Proc. AIAA Meet. Shuttle Environment and Operations*, Washington (1982).
12. Schwartz, S. E. & Johnston, H. S. *J. chem. Phys.* **51**, 1286-1302 (1969).
13. Mende, S. B. et al. *J. Space Rock.* **23**, 189-193 (1986).
14. Korb, L. J., Morant, C. A., Calland, R. M. & Thatcher, C. S. *Bull. Am. ceram. Soc.* **60**, 1188-1193 (1981).
15. NASA Handbook, *Space Shuttle System Payload Accommodations* Vol. XIV (Rev. H, JSC-07700, 1983).
16. Engebretson, M. J. & Mauersberger, K. *J. geophys. Res.* **84**, 839-844 (1979).
17. Glasser, R. P. H. *An Introduction to Chemisorption and Catalysis by Metals* 190-197 (Oxford University Press, New York, 1985).
18. Kofsky, I. & Barrett, J. L. *Planet. Space Sci.* (in the press).
19. Tully, J. C. *J. Chem. Phys.* **73**, 6333-6342 (1980).
20. Köhnlein, W. *Planet. Space Sci.* **28**, 225-243 (1980).

Atmospheric CH_4 , CO and OH from 1860 to 1985

Anne M. Thompson

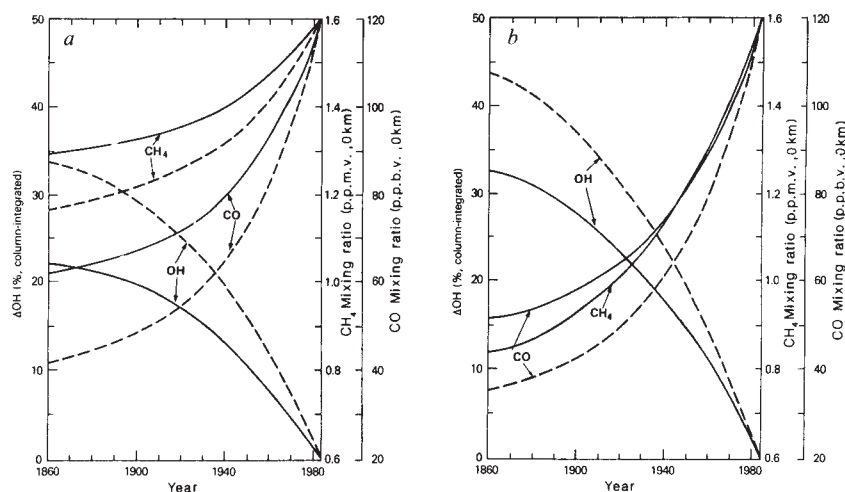
Applied Research Corporation, Landover, Maryland 20785, USA

Ralph J. Cicerone

National Center for Atmospheric Research, Boulder, Colorado 80307, USA

Atmospheric methane, CO and the gaseous OH radical are interdependent: if CH_4 , CO or OH is perturbed, background concentrations of the other two constituents are affected. Perturbations to OH alter photo-oxidation rates of numerous natural and anthropogenic trace gases and affect lifetimes of those species that pass from the Earth's surface to the free troposphere and stratosphere. It is now known that global atmospheric methane con-

Fig. 1 Calculated ground-level methane and CO mixing ratios and per cent changes in column-integrated tropospheric OH from 1860 to 1985. Model conditions are appropriate for the mid-altitude Northern Hemisphere with low background NO_x (20–25 p.p.t.v.). Calculations for low-latitude conditions (higher photolysis rates, lower O_3 and CO mixing ratios) show trends of similar magnitude, implying that the OH changes illustrated here are typical of unpolluted environments. Calculations with higher NO_x levels also show similar trends, but the magnitude of OH change is about half the increase shown here. *a*, Results from models 1 and 2 (Table 1). Methane flux is constant at all times; the calculated CH_4 increase is due solely to CO and OH changes. The solid line is from model runs with a largely anthropogenic upward flux of CO at 35% of the CO source in 1985; the dashed line assumes flux_{CO} (1985) produces 65% of present-day CO. In both cases additional CO derives from CH_4 and C_2H_6 oxidation by OH and from an explicit source, S_s , representing non-methane hydrocarbon oxidation. flux_{CO} and S_s vary in time as described in Table 1. *b*, Results from models 3 and 4 (Table 1). Fixed CH_4 mixing ratios follow ref. 13. Solid and broken lines as in *a*.



centrations are increasing^{1–6}; less definite data suggest that carbon monoxide is also increasing^{1,7–9}. Even before the measurements of refs 1–9 were made, modelling studies of CH_4 –CO–OH coupling had led to predictions^{10–12} of future temporal increases of CH_4 and CO. Here we look backwards in time, using a photochemical model to simulate the trace-gas composition of the unpolluted troposphere at the start of the industrial era (taken as 1860), and at intervals up to 1985. We find that the OH concentration in the background troposphere has decreased significantly and O_3 has increased due to increases of CH_4 and CO; calculated changes depend on temporal trends of NO_x ($\text{NO}_x = \text{NO} + \text{NO}_2$), for which no historical data are available. The calculations allow recent trace-gas trends affecting background chemistry and climate to be viewed in a longer-term context.

In modelling the photochemistry of the pre-industrial atmosphere it is essential to specify past concentrations of NO_x and the more stable species CH_4 , CO and O_3 . Before ~1960, few ambient measurements were made of these trace species, so the input to our model is based on current mixing ratios or source flux estimates extrapolated backwards in time. A summary of key species follows.

CH_4 and CO. We take two extreme methane cases: one with a minimum difference in CH_4 mixing ratio between 1860 and 1985 and one with a maximum difference. Similarly, we assume minimum and maximum temporal changes in CO and combine alternate sets of CH_4 and CO constraints to obtain four model types (Table 1). Models 1 and 2 assume a uniform CH_4 flux (constant sources) at all times, yielding 1.2–1.3 p.p.m.v. (parts per 10^6 by volume) in 1860 and 1.60 p.p.m.v. in 1985 in the models. Models 3 and 4 are based on the assumption that large differences between modern CH_4 and that extracted from ice cores^{13–15} represent real changes in atmospheric composition over the past hundred years, rather than diffusive losses of trapped gases⁵. This is consistent with recently published spectroscopic measurements of past CH_4 (ref. 6). Thus, in models 3 and 4, CH_4 is present at a concentration of only 0.82 p.p.m.v. at ground level in 1860.

Choosing model input for CO is complicated because ground-level CO concentrations vary greatly, owing to the short lifetime of CO and uneven geographical distribution of CO sources^{7,8}. We take 120 p.p.b.v. (parts per 10^9 by volume) CO at ground level as typical for the 1985 mid-latitude Northern Hemisphere; Southern Hemisphere CO concentrations are lower because there are fewer local sources^{16,17}. Uncertainties in CO sources are treated by parameterizing two components in the model

representation of the total source, $S^{\text{total,CO}}$: S_s (a surrogate for the oxidation of non-methane hydrocarbons besides C_2H_6) and flux_{CO} (representing directly injected surface sources). Both S_s and flux_{CO} include natural and anthropogenic contributions as described in Table 1. In ascribing 82% of present-day S_s to natural hydrocarbon sources, we follow Logan *et al.*¹⁷, recognizing that this apportionment is still uncertain. Model runs for years before 1985 assume that natural sources contributing to S_s and flux_{CO} are unchanged and anthropogenic ones scale with population or fossil-fuel usage¹⁸. Thus, S_s changes little in time and flux_{CO} is strongly time-dependent.

Table 1 Model characteristics

Input	Model 1	Model 2	Model 3	Model 4
Past CH_4	High CH_4 : constant 1985 flux		CH_4 specified from ice-core data ¹³	
Past CO	Type 1-CO	Type 2-CO	Type 1-CO	Type 2-CO
Output				
1985 O_3 (0 km)	30.0 p.p.b.v.	29.8 p.p.b.v.	30.0 p.p.b.v.	29.8 p.p.b.v.
1860 O_3 (0 km)	27.7 p.p.b.v.	26.4 p.p.b.v.	25.4 p.p.b.v.	27.1 p.p.b.v.

Type 1-CO: 35% of $S^{\text{total,CO}}$ (1985) is flux_{CO}; type 2-CO: 65% of $S^{\text{total,CO}}$ (1985) is flux_{CO}. The total CO source in year t is:

$$S^{\text{total,CO}}(t) = Q_{\text{CO}}(t) + \text{flux}_{\text{CO}}(t) + S_s(t)$$

where Q_{CO} , the *in situ* photochemical formation from CH_4 and C_2H_6 oxidation, is derived from the model [$Q_{\text{CO}}(z, t)$, integrated from $z = 0$ to 15 km]; S_s (also column-integrated) explicitly represents photo-oxidation of non-methane hydrocarbons (NMHC) besides C_2H_6 :

$$S_s(z, t) = S_s(z, 1985) [f_{\text{NMHC,N}} + r^{\text{ff}}(t) f_{\text{NMHC,A}}(1985)]$$

where $f_{\text{NMHC,N}}$ (0.82) and $f_{\text{NMHC,A}}$ (0.18 in 1985) are fractions of S_s due to oxidation of natural and anthropogenic NMHC¹⁷. $r^{\text{ff}}(t)$ is the ratio of fossil-fuel usage¹⁸ in year t relative to 1985. $S_s(z, t)$ is proportional to the product of the OH number density and an exponential function with the latitude dependence of a typical OH + alkane reaction. flux_{CO}(t), the model lower boundary condition for CO, signifies direct surface-injected CO sources:

$$\text{flux}_{\text{CO}}(t) = \text{flux}_{\text{CO}}(1985) [f_{\text{flux,N}} + r^{\text{ff}}(t) f_{\text{flux,A}}^{\text{ff}}(1985) + r^{\text{pop}}(t) f_{\text{flux,A}}^{\text{ff}}(1985)]$$

where $r^{\text{pop}}(t)$ is the ratio of the population in year t relative to that in 1985; f values are based on a budget¹⁷ for Northern Hemisphere CO sources; $f_{\text{flux,N}} = 0.13$ is the natural fraction of flux_{CO} (oceanic flux, emissions from plants and lightning-induced wildfires); $f_{\text{flux,A}}^{\text{ff}}$ and $f_{\text{flux,A}}^{\text{ff}}$ are fossil-fuel and non-fossil-fuel (wood fuel and biomass burning) combustion components of anthropogenic flux_{CO}. In all models C_2H_6 at 0 km was increased from 1.0 to 1.5 p.p.b.v. between 1860 and 1985. This has negligible effect on OH and CO changes.

O₃. Ozone concentrations vary so much in time and space, even on a short timescale, that a realistic determination of tropospheric O₃ in 1860 is impossible. For simplicity we assume identical boundary conditions for O₃ in current and past-year models: these conditions are fixed stratospheric input and fixed deposition velocity¹⁹.

H₂O, NO_x. These species and other model parameters, for example ultraviolet and visible radiation and rainout rates, are critical in determining O₃ and OH concentrations. Water vapour, radiation fluxes and rainout rates for soluble species are fixed at constant levels¹⁹ in all model runs. Boundary conditions for odd nitrogen (NO_x) are assumed to be uniform at all times; a stratospheric odd-nitrogen source controls NO_x in the upper and mid-troposphere and a small upward flux of NO gives 20–25 p.p.t.v. (parts per 10¹² by volume) NO_x in the boundary layer. Although NO_x sources include a large anthropogenic component that has increased during the past century, it seems reasonable to suppose that the low levels of NO_x that determine O₃ and OH in the background troposphere are supplied by non-anthropogenic sources^{20,21} (for example, stratospheric injection, lightning, weak oceanic and soil fluxes of NO) which have not changed appreciably with time.

Details of the model used in the present paper and profiles of tropospheric O₃, NO_x and OH are reported in ref. 19. The model solves for a standard complement of odd O, H, N species; CO, CH₄ and species derived from CH₄ and C₂H₆ oxidation, including peroxyacetylnitrate (PAN). Simulations are performed with a steady-state model at 10–30-yr intervals. CH₄ and CO can exhibit slower transients than would be predicted from their respective 10-yr and 2-month lifetimes¹¹, depending on the mix of CO sources.

Figure 1 shows ground-level CH₄ and CO mixing ratios and changes in column-integrated OH (0–15 km) from 1860 to 1985. Figure 1a shows results assuming minimum CH₄ change between 1860 and 1985 (models 1 and 2), and Fig. 1b assumes maximum CH₄ change (models 3 and 4). Tropospheric OH decreases as CO and CH₄ increase, with the greatest change occurring when both CO and CH₄ undergo maximum change (model 4). In that case CH₄, CO and fractional OH increases in 1950 (relative to 1985) are similar to those presented by Levine *et al.*²², although our NO_x level is much lower. When we calculate perturbed CH₄–CO–OH cycles at higher NO_x levels (1 p.p.b.v. at 0 km) we find that increased CH₄ and CO cause near-surface OH to increase, but total tropospheric OH decreases as in Fig. 1 (although by only half as much).

Changes in atmospheric lifetimes of hundreds of trace gases, for example, reactive hydrocarbons, sulphur and nitrogen compounds, may have occurred since 1860 as a consequence of OH perturbations. Figure 1a (models 1 and 2), for which no change in surface CH₄ emissions is assumed, shows that CO increases and OH decreases could have caused background-level CH₄ to increase by 24–37% from 1860 to 1985. Figure 1b, however, implies that CH₄ doubling since 1860 requires a change in CH₄ fluxes as well as altered background OH in response to CO changes.

Changes in CO–CH₄–OH chemistry affect climate because O₃, CH₄ and CO₂ (to which CO converts by OH oxidation) are 'greenhouse' gases. Increases in CH₄ and CO cause tropospheric ozone to increase because CH₄ and CO contribute to photochemical O₃ production^{23–25}. The temporal changes shown in Fig. 1 imply an 8–18% tropospheric O₃ increase from 1860 to the present (Table 1); observations point to a recent temporal increase in tropospheric O₃ (refs 26, 27). O₃ perturbations instantaneously alter the Earth's infrared radiative forcing. If there has been sufficient time for a temperature response, this could have caused a small rise in global temperature, depending on the altitude profile of the O₃ increase^{28,29}. Atmospheric CO increases affect radiation by supplying CO₂ (provided that OH remains at adequate levels), and by suppressing OH and allowing O₃ and CH₄ to accumulate, even if CH₄ emissions do not

increase. For example, atmospheric CO₂ formed from the reaction OH+CO in 1860 (8.16×10^{10} molecules cm⁻² s⁻¹, integrated through the troposphere) exceeded the direct fossil-fuel source of CO₂, that is, 2.6×10^{10} molecules cm⁻² s⁻¹ (ref. 18). Approximately 75% of this CO was natural in origin (Table 1).

A.M.T. acknowledges partial support from the NASA Tropospheric Chemistry Office. The National Center for Atmospheric Research is sponsored by NSF.

Received 25 November 1985; accepted 17 February 1986.

1. Graedel, T. E. & McRae, J. E. *Geophys. Res. Lett.* **7**, 977–979 (1980).
2. Rasmussen, R. A. & Khalil, M. A. K. *J. geophys. Res.* **86**, 9826–9832 (1981).
3. Blake, D. R. *et al. Geophys. Res. Lett.* **9**, 477–480 (1982).
4. Rasmussen, R. A. & Khalil, M. A. K. *J. geophys. Res.* **89**, 11599–11605 (1984).
5. Ehhalt, D. H., Zander, R. J. & Lamontagne, R. A. *J. geophys. Res.* **88**, 8442–8446 (1983).
6. Rinsland, C. P., Levine, J. S. & Miles, T. *Nature* **318**, 245–249 (1985).
7. Seiler, W., Giehl, H., Brunke, E. & Halliday, E. *Tellus* **36B**, 219–231 (1984).
8. Khalil, M. A. K. & Rasmussen, R. A. *Science* **224**, 54–56 (1984).
9. Rinsland, C. P. & Levine, J. S. *Nature* **318**, 250–254 (1985).
10. Chameides, W. L., Liu, S. C. & Cicerone, R. J. *J. geophys. Res.* **82**, 1795–1798 (1977).
11. Sze, N. D. *Science* **195**, 673–674 (1977).
12. Hameed, S., Pinto, J. P. & Stewart, R. W. *J. geophys. Res.* **84**, 763–768 (1979).
13. Khalil, M. A. K. & Rasmussen, R. A. *Atmos. Environ.* **19**, 397–407 (1985).
14. Stauffer, B., Fischer, G., Neftel, A. & Oeschger, H. *Science* **229**, 1386–1388 (1985).
15. Craig, H. & Chou, C. C. *Geophys. Res. Lett.* **9**, 1221–1224 (1982).
16. Seiler, W. *Tellus* **26**, 116–135 (1974).
17. Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
18. Keeling, C. D. *Tellus* **25**, 174–198 (1973).
19. Thompson, A. M. & Cicerone, R. J. *J. geophys. Res.* (submitted); **87**, 8811–8826 (1982).
20. Levy II, H., Mahlman, J. D. & Moxim, W. J. *Geophys. Res. Lett.* **7**, 441–444 (1980).
21. Liu, S. C., McFarland, M., Kley, D. & Zafriou, O. C. *J. geophys. Res.* **88**, 1360–1368 (1983).
22. Levine, J. S., Rinsland, C. P. & Tennille, G. M. *Nature* **318**, 254–257 (1985).
23. Fishman, J., Solomon, S. & Crutzen, P. J. *Tellus* **31**, 432–446 (1979).
24. Fishman, J. & Seiler, W. *J. geophys. Res.* **88**, 3662–3670 (1983).
25. Chameides, W. L. & Walker, J. C. G. *J. geophys. Res.* **78**, 8751–8760 (1973).
26. Bojkov, R. *Spec. Envir. Rep.* no. 16 (World Meteorological Organization, Geneva, 1984).
27. Logan, J. A. *J. geophys. Res.* **90**, 10463–10482 (1985).
28. Fishman, J., Ramanathan, V., Crutzen, P. J. & Liu, S. C. *Nature* **282**, 818–820 (1979).
29. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).

Spreading direction in the central South China Sea

Guy Pautot*, Claude Rangin†, Anne Briais‡, Paul Tapponnier‡, Paul Beuzart*, Gilles Lericolais*, Xavier Mathieu§, Jinlong Wu||, Shuqiao Han||, Hengxiu Lil||, Yingxian Lu|| & Jicheng Zhao||

* IFREMER, Centre de Brest, BP 337, 29273 Brest Cedex, France

† Université de Paris 6, Département de Géologie Structurale, 4 Place Jussieu, 75252 Paris Cedex 05, France and Institut Français du Pétrole, l'Avenue de Bois-Préau, 92500 Reuil Malmaison, France

‡ Institut de Physique du Globe, Laboratoire de Tectonique et Mécanique de la Lithosphère, Université de Paris 6, 4 Place Jussieu, 75252 Paris Cedex 05, France

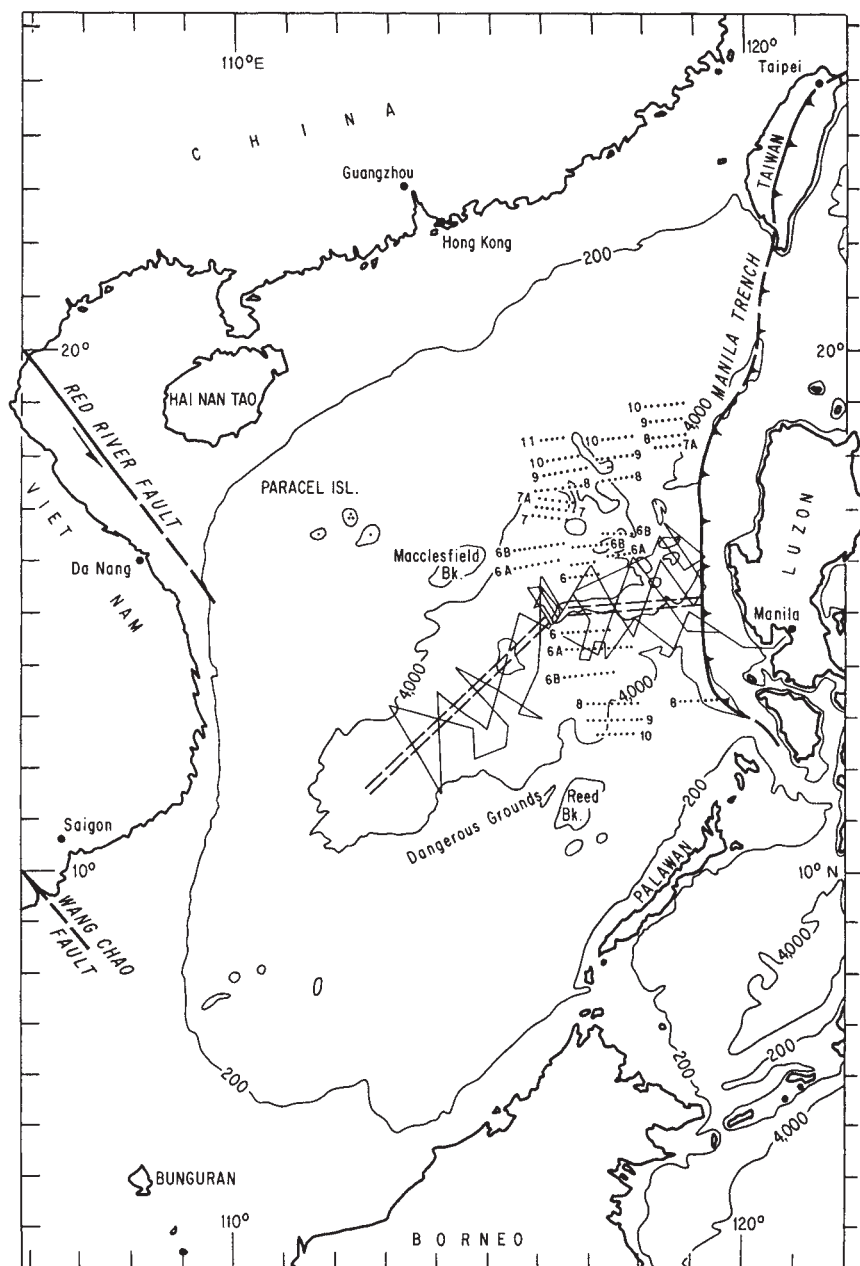
§ Université de Bretagne Occidentale, 615 "Océanologie et Géodynamique" 6, Avenue Le Gorgeu, 29283 Brest Cedex, France

|| First Institute of Oceanography, National Bureau of Oceanography, PO Box 98, Qingdao, China

Recent work^{1,2} indicates that the South China Sea is an 'Atlantic-type' marginal basin of late Tertiary age. Magnetic anomalies in the eastern part of the sea are consistent with seafloor spreading directed approximately north-south^{1,2}. We present here a new morphostructural study based on coupled seabeam mapping and single-channel seismic reflection profiling, which reveals dominant normal fault scarps, striking N50° E between 113 and 119° E longitude near the axis of this basin. Such a structural fabric implies a NW–SE spreading direction, at least in the 150–200-km-wide axial region of the South China Sea, and places new constraints on geodynamic models for the formation of this basin in the tectonic and palaeogeographic framework of South-East Asia and the South-West Pacific.

The South China Sea (Nanhai), largest of the marginal basins of the western Pacific, is bounded by the continental margins

Fig. 1 Bathymetry of the South China Sea. Tracks of RV *Jean Charcot* during the Nanhai survey are reported. Also shown are major structures such as the Red River Fault and Manila Trench. Dashed parallel lines indicate inferred position of ridge axis^{1,2}. Magnetic anomalies are from ref. 2.



of South China, Vietnam and Borneo, and by the Manila Trench, where its crust is now being consumed by eastward-dipping subduction (Fig. 1). East of 115° E, its ~ 700 -km-wide oceanic floor, where east-west-trending magnetic anomalies 11 to 5e have been identified¹⁻³, appears to have formed mostly between 32 and 17 Myr (refs 1, 2). The east-west-trending Scarborough Seamounts probably mark the location of the spreading axis in this part of the basin^{1,2}. West of 115° E, the abyssal plain divides into two wedge-shaped basins which narrow westwards, respectively north and south of the Macclesfield-Paracels high¹. Magnetic anomalies have not yet been identified in these basins, although the deeper crust appears to be of oceanic nature¹, and the axis of the larger, southern basin, trending $N50^{\circ}$ E, is marked by a -30 -mgal gravity low^{1,2} (Fig. 1). The difference between the latter NE-SW structural trend and the more nearly east-west orientation of magnetic anomalies east of 115° E has been the source of conflicting interpretations. While Bowin *et al.*³ inferred a roughly NE-SW fabric extending into the entire eastern basin, other reconstructions^{1,2,4,5} imply oblique, north-south spreading in the southwestern sub-basin. Alternatively, the complexity of

the South China basin could result from two (or more) distinct phases of spreading⁴.

In order to obtain a clearer view of the morphology and structure of the axial floor of the South China Sea, RV *Jean Charcot* performed, between 113 and 119° E, a survey comprising 74 profiles in which seabed mapping was coupled with single-channel seismic, magnetic and gravity recording (Figs 1, 3). The combination of seabed with seismic reflection allowed the correlation of faults identified on the seismic sections with morphological scarps outlined by steep gradients or rectilinear bathymetric contours on the sea floor (Fig. 4). The width of the seabed swaths (2 – 3 km in most of the survey area) allowed the direct determination of fault orientations, so that the kinematics of spreading could be constrained more tightly than in previous studies¹⁻³.

Major normal faults observed on seismic profiles (Figs 3, 4) are shown in Fig. 2, with the corresponding seabed-derived scarp orientations. The rather thin sedimentary cover near the spreading axis does not hide the morphological expression of the faults (Figs 3, 4). Along most of the length of the ridge, we

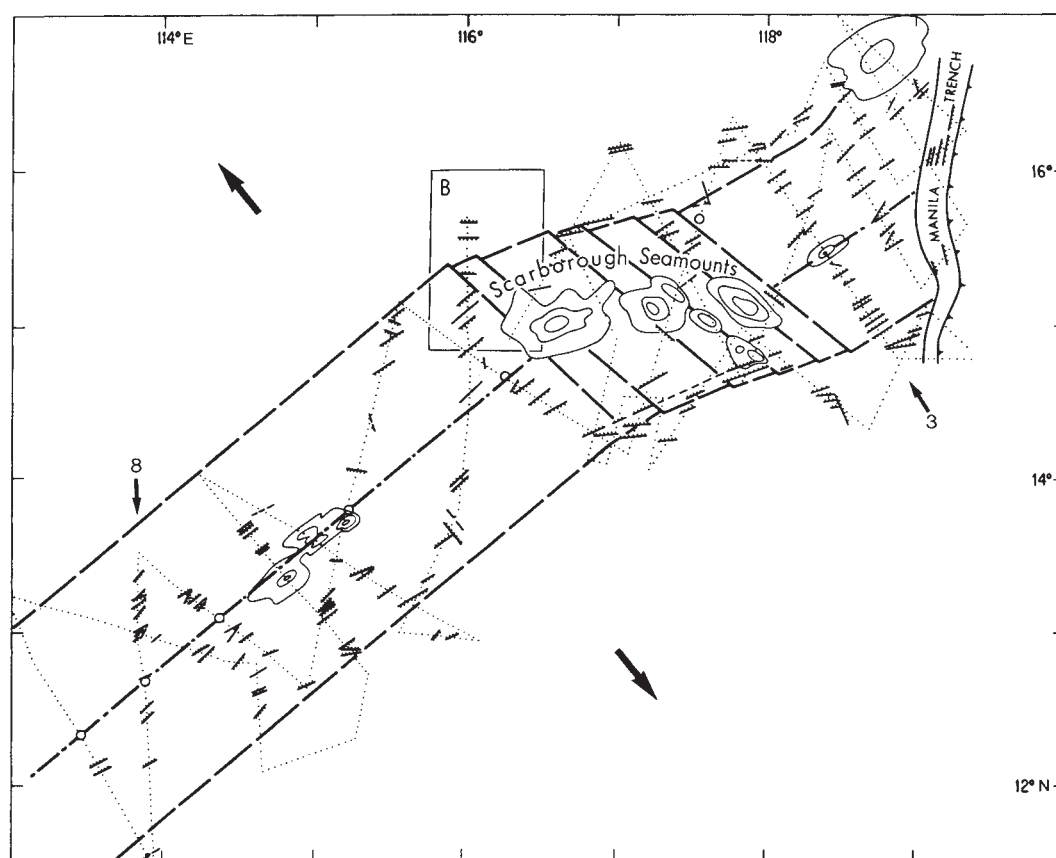


Fig. 2 Strikes and dips of major faults derived from combined analysis of seabeam and seismic reflection records. Dotted lines represent RV *Jean Charcot* profiles. Profiles 3 and 8 (Fig. 3) are indicated. Box B marks an area with denser seabeam coverage. Thin contours represent major seamounts (Scarborough chain). Thick dashed lines indicate outer limit of zone with predominant N50° E-trending scarps, and inferred transverse fault zones between 116 and 118° E. Dash-dot line outlines probable location of spreading axis.

can identify the spreading axis itself by the inwards-facing of the fault scarps and the outwards-tilting of blocks. As shown in Fig. 2, most of the normal faults strike $N50 \pm 10^\circ$ throughout the axial region of the South China Sea, implying that the most recent spreading direction was roughly orthogonal ($N140^\circ$ E). Outside the central 150 km, scarps oriented $N70$ – 80° E are found, particularly between 116 and 119° E, north and south of the Scarborough Seamounts (Fig. 2).

At a more detailed level, it is possible to subdivide the axial region of the South China Sea into three distinct segments:

(1) A southwestern segment, which extends linearly from about 116° E towards the south-west, with a width of at least 150 km. The rather smooth oceanic acoustic basement, with outward-tilted blocks, is here covered by 1–2 s (seismic two-way travel time) of sediments. To the south, it appears to be bounded by structural highs trending $N50$ – 70° E, characterized by deeper reflectors, which may correspond to faulted blocks of attenuated continental crust (Fig. 3, profile 8).

(2) A northeastern segment, between about 118° E and the Manila Trench, with a NE–SW trend comparable to that of the southwestern segment. Inward-facing fault scarps striking $N50^\circ$ E are particularly clear up to ~ 75 km from the axis (Fig. 2), which is here marked by a gravity low. Farther away from the axis, major fault scarps still face mostly inwards but have more varied orientations, predominantly $N70$ – 80° E. To the north, near the Manila Trench, a large seamount masks the area in which fault scarp orientations may change.

(3) A central segment, between about 116 and 118° E, with a more complex structure and morphology. Many of the seamounts of the Scarborough chain have formed along the axis of this segment. These large submarine volcanoes dominate the fabric of the sea floor in the axial region. In addition, the

sediment cover in this area is thicker than in the adjacent segments. Despite the coating of the sea floor by sediments and late volcanic products, however, scarps trending $N50$ – 60° E are observed close to the inferred axis. They appear to be dissected by scarps trending $N125$ – 140° E which are more numerous in this segment of the ridge than elsewhere. Moreover, the Scarborough Seamounts and minor volcanic constructions near them are elongated in these two nearly orthogonal directions (particularly the NW–SE one), which thus appear to have controlled the ascent of late basaltic volcanism. This $N50^\circ$ E, $N135^\circ$ E seafloor structural fabric is especially clear in the area near $15^\circ 30'$ N, 116° E in which a dense seabeam coverage was obtained (box B in Fig. 2). At distances on the order of 80 km from the axis, scarps trending $N70$ – 80° E are observed.

The *Jean Charcot* survey provides new insights into the latest stages of spreading in the South China Sea. Between 113 and 119° E, the 150–200-km wide axial region of the ridge appears to be composed of two rather linear, continuous, north-east-trending segments, separated by a central, east-west-trending segment where the predominant fabric of normal faults striking $N50^\circ$ E and bounding outward-tilted blocks is disrupted by transverse faults striking $N135^\circ$ E, and where the axis has been the site of late intrusions along these dominant fault trends.

Both the small- and large-scale geometries suggest that seafloor spreading took place throughout the axial region in a NW–SE direction. The offset of the northeastern and southwestern segments of the ridge, and the presence of NW–SE-striking fault scarps and elongated volcanic highs in the central segment is easily reconciled with the above spreading direction if such scarps and highs are interpreted to mark left-lateral transform fault zones. Progressive left-lateral displacement of the spreading axis across such transform zones would thus

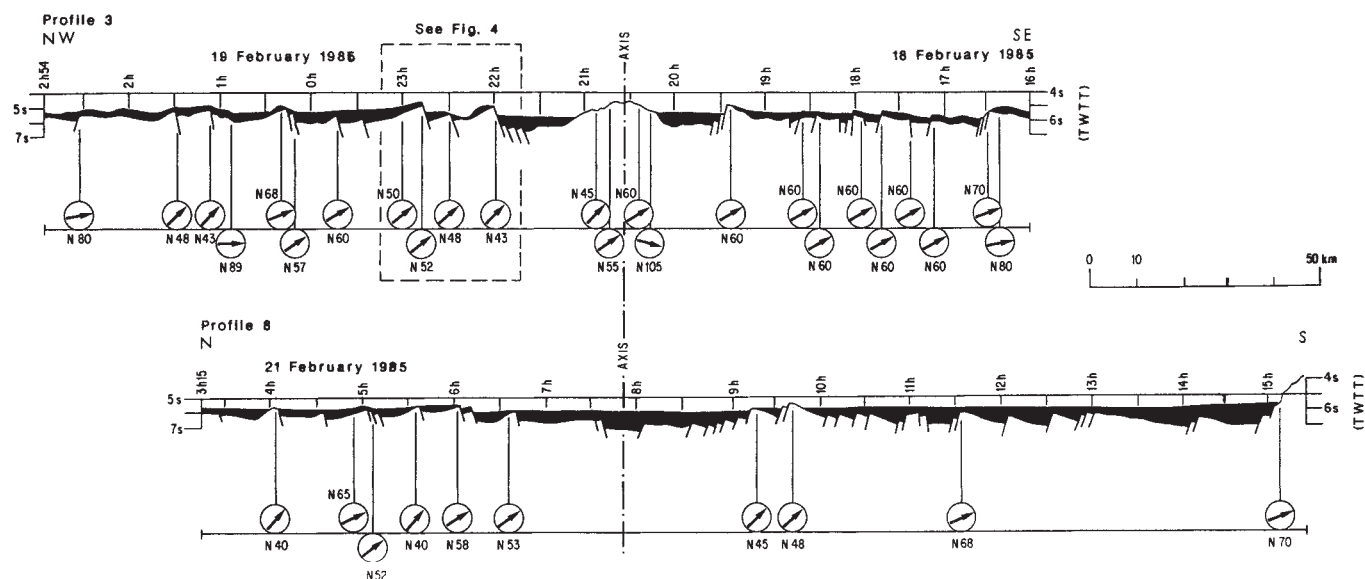
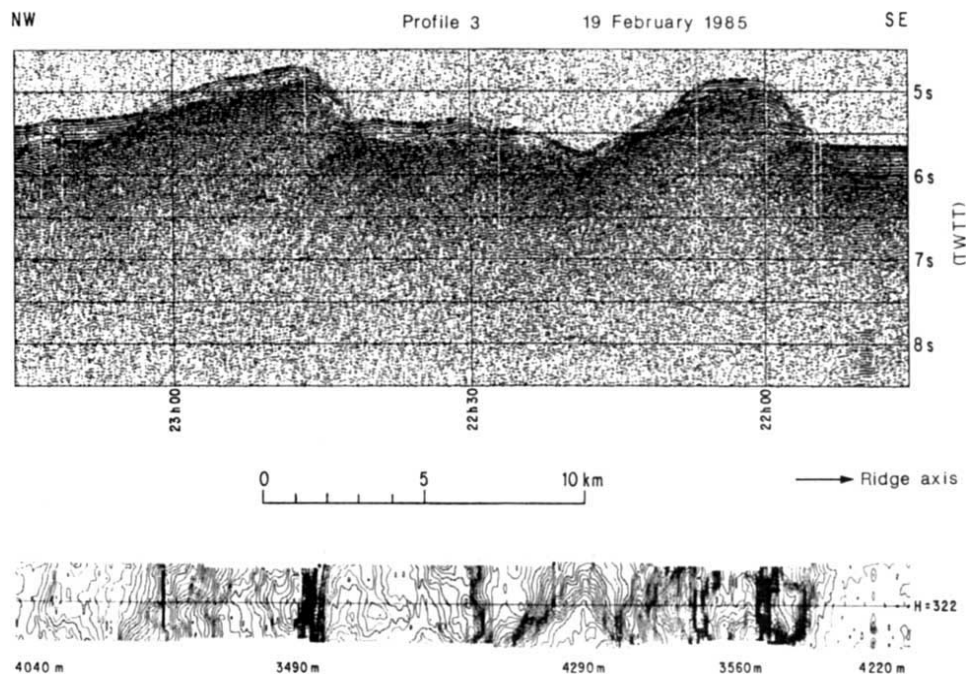


Fig. 3 Schematic sections drawn from single-channel seismic profiles 3 and 8 (sediment cover in black), and orientation of fault scarps (arrows) seen on corresponding seabed records. TWTT, two-way travel time.

Fig. 4 Coupled single-channel seismic reflection profile (vertical exaggeration = 3) and seabed record along part of profile 3 (box in Fig. 3). Horizontal scale is the same. Note clear northwestward-tilted blocks and southeastward-dipping normal faults. Steep gradients in bathymetric contours outline fault scarps striking N43–52° E (Fig. 3).



account for the roughly east-west average trend of the central ridge segment. This more densely fractured, and hence particularly weak crustal zone would have been a favoured injection site for basaltic magmas, allowing the formation of the Scarborough seamount chain which now outlines the ridge axis between 116 and 118° E at 15° N (Fig. 2).

The *Jean Charcot* 'Nanhai' survey discussed here was mostly confined within the area bounded by magnetic anomaly 6a^{1,2} (Fig. 1), which, along with anomaly 6, was clearly identifiable on profiles made near 116° E, 16° N during a second survey to the north of the axial zone (A.B. and P.T., in preparation). Therefore, in a strict sense, the geometry and kinematics outlined above concern only the sea floor generated after about 20 Myr (that is in the most recent 3–4 Myr of spreading^{1,2,4}), at a rate comparable to that estimated by Taylor and Hayes² (~5 cm yr⁻¹). Although the NW-SE spreading direction inferred here appears to remain approximately constant along strike, the youngest oceanic floor may have formed in two distinct, crustal

environments, east and west of 115° E. West of 115° E this floor was perhaps emplaced within stretched continental crust^{1,2,4}, the average orientation of fault scarps and faulted blocks north of the southwestern inner basin being apparently compatible with a similar NW-SE extensional regime (A.B. and P.T., in preparation). East of 115° E, on the other hand, the inner oceanic floor formed in the centre of a 500-km-wide area already floored by oceanic crust, perhaps with a different structural fabric¹⁻³.

The difference between the interpretation presented here and those of previous workers¹⁻⁴ follows mostly from the first use of the seabed imaging technique in exploring the floor of the South China Sea. In establishing the identity of a NW-SE spreading regime in the entire axial region of the sea, this interpretation is in keeping with that of Bowin *et al.*³, and appears to contradict the north-south direction of spreading inferred by Taylor and Hayes^{1,2} from nearly east-west-trending magnetic anomalies, all the way into the axis of the eastern basin. Since such anomalies are clearest outside the area of our

seabeam survey, however, it is now important to determine whether the pattern of faulting observed near the axis extends outside the 150–200-km inner part of the eastern basin. The presence of N60–80° E scarps in the outermost sections of the seabeam profiles between 116 and 119° E (Fig. 2) suggests that it may not. On the other hand, seabeam profiles made well north of the 200-km-wide axial zone, albeit west of 116°30' E during a shorter survey following 'Nanhai' (A.B. and P.T., in preparation) also reveal N50° E-striking normal fault scarps.

A change from N50° E to N80° E in the strikes of normal faults ~100 km away from the axis could indicate a major anti-clockwise rotation of the spreading direction ~20 Myr ago. Such a major kinematic change appears to be noticeable on land in Taiwan and in the Philippines, where incipient subduction along the Manila Trench is apparently marked by new (late Oligocene–Lower Miocene) volcanism and intrusion in the western cordillera of Luzon^{6,7}, immediately followed by obduction of middle Oligocene ophiolites in Mindoro⁸. That change could represent the response of the spreading centre to new kinematic constraints along the nearest plate boundary and resembles that documented at about the same time within the Shikoku Basin (N. Chamot-Rooke, V. Renard and X. Le Pichon, in preparation).

Conversely, as suggested by close comparison of Figs 1 and 2, the orientation of the magnetic isochrons (N80° E) could be distinct from that of normal faulting (N50° E) in the inner as well as the outer part of the eastern South China Sea. One way to reconcile magnetic anomalies trending N80° E with normal fault scarps striking N50° E and a NW–SE spreading direction is implicit in the geometry of the disrupted central ridge segment in Fig. 2: magnetic isochrons may smooth out left-lateral offsets on closely spaced NW–SE transform faults (A.B. and P.T., in

preparation). In this event, the near-parallelism between the NW–SE spreading direction, the left-lateral transform faults in the central South China Sea, and the large, left-lateral, Tertiary strike-slip faults of South China and Sundaland (Red River, Wang Chao, Three Pagodas)^{9,10} could imply, in keeping with the tenets of 'rigid' plate tectonics, that the opening of that sea was driven by the southeastwards extrusion, forced by the India–Asia collision, of different parts of Sundaland along those faults^{9–12}.

This survey was carried out within a Cooperative Program (Nanhai project) between the National Bureau of the Sea (Beijing) and IFREMER (Paris). We thank all participants, particularly Commandant Hubert Guidal and the crew, for collaboration on board RV *Jean Charcot*. This is IFREMER contribution 65 and IGP contribution 902.

Received 31 January; accepted 24 February 1986.

1. Taylor, B. & Hayes, D. E. in *The Tectonic and Geologic Evolution of Southeast Asian Seas and Islands* Pt 1 (ed. Hayes, D. E.) (American Geophysical Union, Washington, DC, 1980).
2. Taylor, B. & Hayes, D. E. in *The Tectonic and Geologic Evolution of Southeast Asian Seas and Islands* Pt 2 (ed. Hayes, D. E.) (American Geophysical Union, Washington, DC, 1983).
3. Bowin, C., Lu, R. S., Lee, C. S. & Schouten, H. *Bull. Am. Ass. Petrol. Geol.* **62**, 1645–1672 (1978).
4. Holloway, N. *Bull. Am. Ass. Petrol. Geol.* **66**, 1355–1383 (1982).
5. Hayes, D. E. *Energy* **10**, 373–382 (1985).
6. Bachman, S. B., Schweller, W. J. & Lewis, S. D. *Bull. Am. Ass. Petrol. Geol.* **67**, 1143–1162 (1982).
7. Wolfe, J. A. *J. geol. Soc. Philipp.* **35**, 1–30 (1981).
8. Rangin, C., Stephan, J. F. & Muller, C. *Geology* **13**, 425–428 (1985).
9. Taponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R. & Cobbold, P. *Geology* **10**, 611–616 (1982).
10. Taponnier, P., Peltzer, G. & Armijo, R. in *Collision Tectonics* (eds Coward, M. P. & Ries, A. C.) 115–157 (Geological Society Special Publication, London, 1986).
11. Zhao, W. L. & Morgan, W. J. *Tectonics* **4**, (4) 359–269 (1985).
12. Peltzer, G., Taponnier, P. & Cobbold, P. *C. r. heb. Séanc. Acad., Sci., Paris* **294**, 1341–1348 (1982).

Propagation of internal tides from the upper slopes of the Bay of Biscay

R. D. Pingree, G. T. Mardell & A. L. New

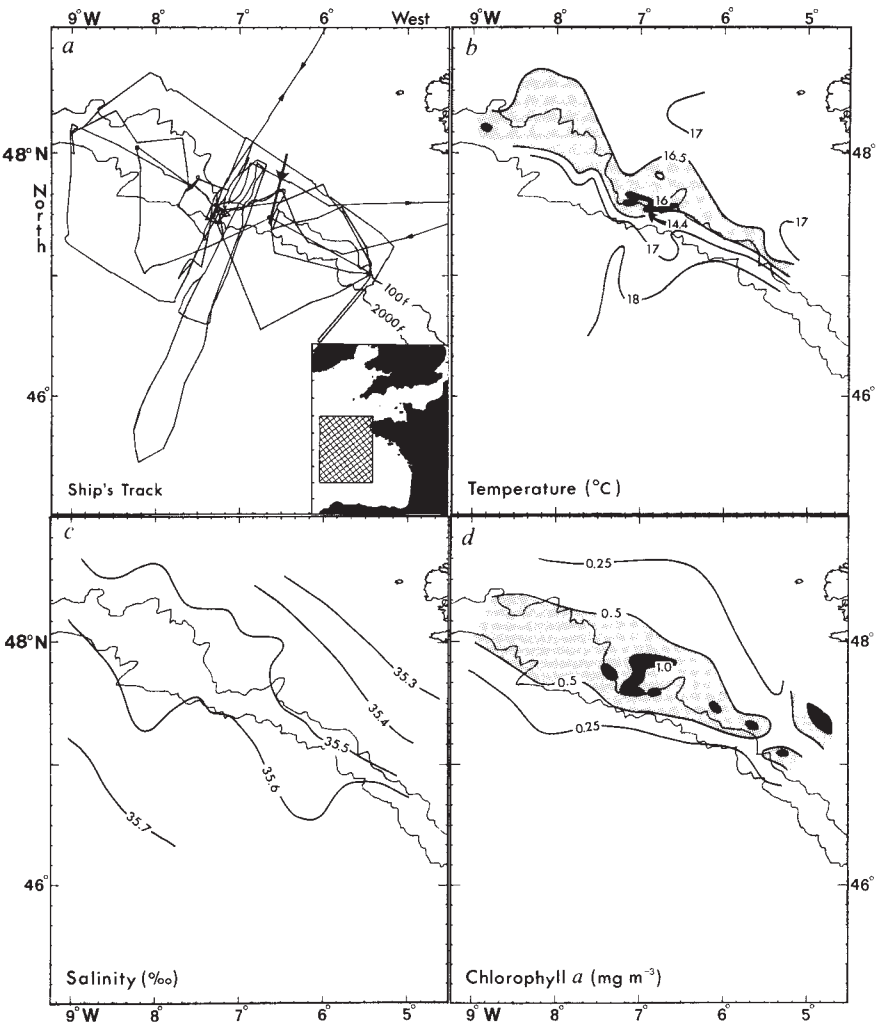
Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK

The shelf-break region of the Celtic Sea (north of the Bay of Biscay) is characterized by a reduced sea-surface temperature (~1 °C) throughout the summer months¹. The cool region stretches for ~300 km along the slopes and exhibits enhanced levels of surface phytoplankton abundance. It has been postulated² that physical mixing creates this favourable environment for phytoplankton growth by mixing inorganic nutrients from below the seasonal thermocline, allowing the plant cells to grow more favourably in the well-illuminated surface waters. One possibility is that the mixing is due to the production and propagation of internal tides^{3–6}. The present study was undertaken to detect and observe the nature of the internal tide at the Celtic Sea shelf break. We show here that internal tides, generated at spring tides close to the 200-m contour^{7–9}, propagate simultaneously both on shelf and into the ocean as distorted progressive waves, and when this happens, nutrients are mixed towards the sea surface.

In 1982 and 1983, the RV *Frederick Russell* studied internal waves and tides propagating on-shelf across the Celtic Sea^{9,10}, whereas in 1984 and 1985 more attention was focused on the off-shelf propagating waves. The September 1985 results were obtained during the participation of the RV *Frederick Russell* in the French *ONDINE* exercise, which examined internal motions and mixing in the stratified surface waters of the Bay of Biscay (to the west of France) and neighbouring shelf regions. The characteristic shelf-break cooling and associated increases in surface levels of chlorophyll *a* (with respect to neighbouring

shelf and oceanic waters) were evident (Fig. 1) and at spring tides repeated runs, in a narrow section covering a 300-km extent normal to the shelf break, were made using the SeaSoar to identify the structure of the internal tide. The SeaSoar is a towed undulating conductivity–temperature–depth/fluorometer system^{11,12} which allows vertical profiles of temperature, salinity and chlorophyll *a* to be obtained while steaming at 8 knots. Some typical examples of on-shelf and off-shelf waves measured with the SeaSoar are shown in Figs 2 and 3. The internal tide (period 12.4 h) is often characterized by structures of higher frequency (period ~20 min relative to a fixed observer) and higher wavenumber (wavelength ~1 km)¹⁰ which appear mainly in the troughs of the internal tide. Some increase in amplitude of the shorter-period waves in this position would be expected due to compressional straining of the higher-wavenumber structures by the internal tide. In addition, the shorter-period internal waves may be generated locally near this position because the leading slope of the internal tidal trough steepens initially as it propagates from the shelf-break region. The more slowly propagating shorter waves may be partially trapped in the internal tidal trough, where a favourable current structure effectively advects them along. The sea-surface signatures of the higher-wavenumber structures could be observed visually, and ship radar observations^{13,14} showed that these waves propagated in a direction (020–200°) normal to the shelf break. The internal tide has an asymmetric structure, where dispersion due to non-linearity is balanced by that due to Earth's rotation⁹, and the characteristic of deeply penetrating troughs together with the associated higher-wavenumber structure has been used to identify all the longer waves encountered on the SeaSoar run. A travel-time graph can be obtained by plotting the position (distance from the 200-m contour) and time (with respect to high water at Plymouth) of each wave. The results are shown in Fig. 4a, where the slope of the line represents the speed of travel of the internal tides from the shelf break. The variable speed across the shelf is caused by the barotropic tidal currents alternately

Fig. 1 *a*, Map of the region showing the cruise track of the RV *Frederick Russell* for the period 7–25 September 1985. Also shown (bold line) is the track and direction for the SeaSoar run (30 August 1984) illustrated in Fig. 2. Dots indicate positions of current-meter moorings. *b*, Mean sea-surface temperature contours for the period 13–25 September 1985. *c*, Mean sea-surface salinity contours (‰) obtained while steaming along the track shown in *a*. *d*, Mean surface levels of chlorophyll *a* (mg m^{-3}) obtained while steaming along the track shown in *a*.



assisting and opposing the on-shelf progress. Apart from an initial assistance by the barotropic tidal currents, the off-shelf propagating tide has a steadier oceanward progress since the barotropic tidal currents are only a few cm s^{-1} in the deeper (4,500 m) water regions. Scatter of the points about an idealized line is to be expected as the waves may be refracted by varying density structure¹⁵, propagate through oceanic eddies^{16–18}, and originate from irregular topography in the upper slope region where there are also variations in the timing of maximum off-shelf tidal streaming in an along-slope direction.

Two interesting points that emerge from the travel graphs

concern the generation times of the internal tidal troughs and the differences in mean phase speed for the on- and off-shelf waves. Both the on- and off-shelf travelling troughs seem to be generated near the 200-m contour at about the same tidal state, that is, ~2–4 h after high water at Plymouth, which corresponds to the off-shelf streaming phase for the barotropic tide (maximum off-shelf tidal streaming in a direction normal to the slope occurs ~2 h after high water at Plymouth in this region), ruling out the possibility that they propagate alternately on-shelf and off-shelf due to some form of simple lee-wave mechanism¹⁹. Although the actual physical mechanism has not been estab-

Table 1 Estimates of phase speeds for internal tides

Internal tide	h_1 (m)	h_2 (m)	$\Delta\rho/\rho \times 10^3$	Phase speeds (m s^{-1})		
				Using equations (1) and (2)	Numerical studies†	Observations
Shelf	48	122	0.74	0.77	Mode	0.70
					1 0.70	
Ocean	55	4,500	1.1	(1.2)*	2 0.26	1.03
					Mode	
					1 3.5	
					2 1.6	
					3 1.2	
					4 0.95	
					5 0.73	

* Equations (1) and (2) not valid for the deep ocean.
† Assumes a flat bottom.

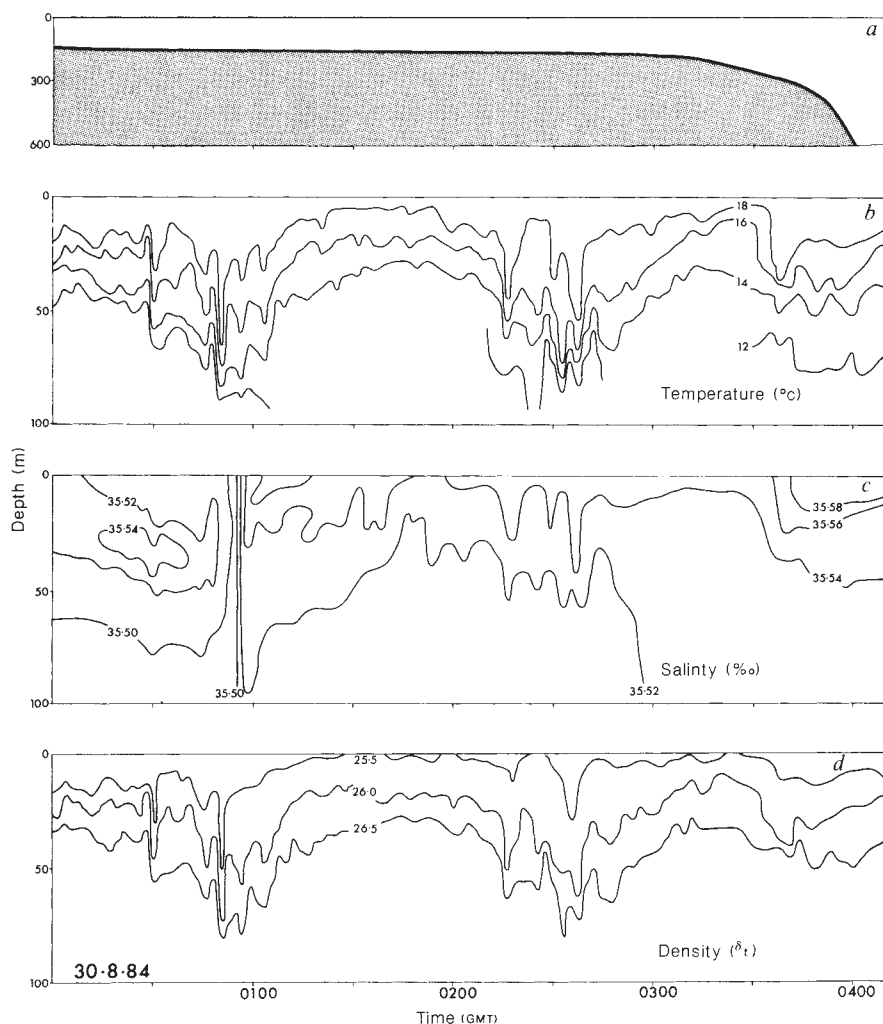


Fig. 2 Simultaneous vertical profiles of water depth (a), temperature (b), salinity (c) and density (d) obtained along the SeaSoar run shown in Fig. 1a. The temperature, salinity and density profiles were contoured from 100 individual profiles and show two deeply penetrating internal tidal troughs travelling on-shelf. Note the asymmetric structure of the long waves and the higher-wavenumber structures in troughs of the internal tide.

Fig. 3 Simultaneous SeaSoar profiles of temperature and chlorophyll *a*, showing two ocean propagating internal tidal troughs. The section was contoured from 180 vertical profiles and the run was made while the ship was heading towards the shelf. The two troughs at ~1300 and 1600 GMT can be identified in Fig. 4a and b as numbers 10 and 11. The total water depth (m) is indicated below the temperature section. The surface chlorophyll *a* increases at the shelf break (200 m), as depicted in Fig. 1d.

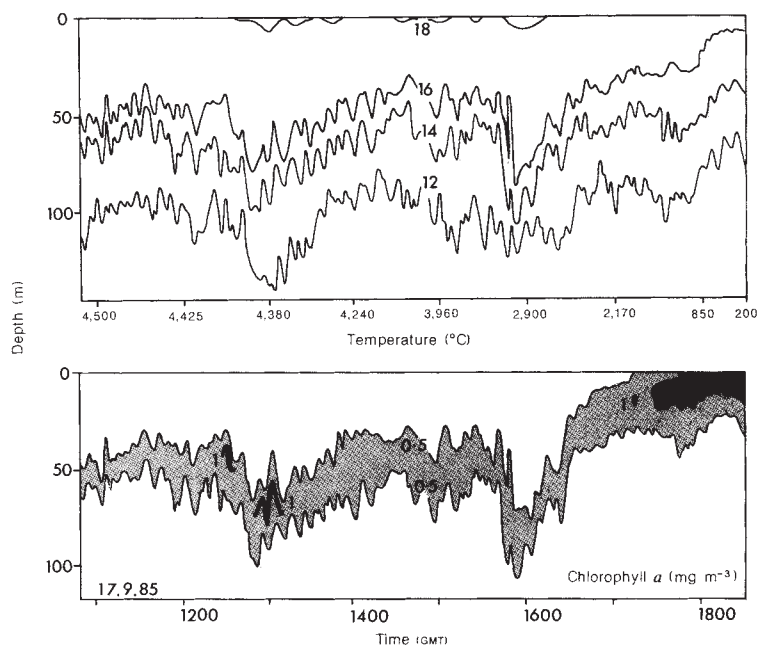
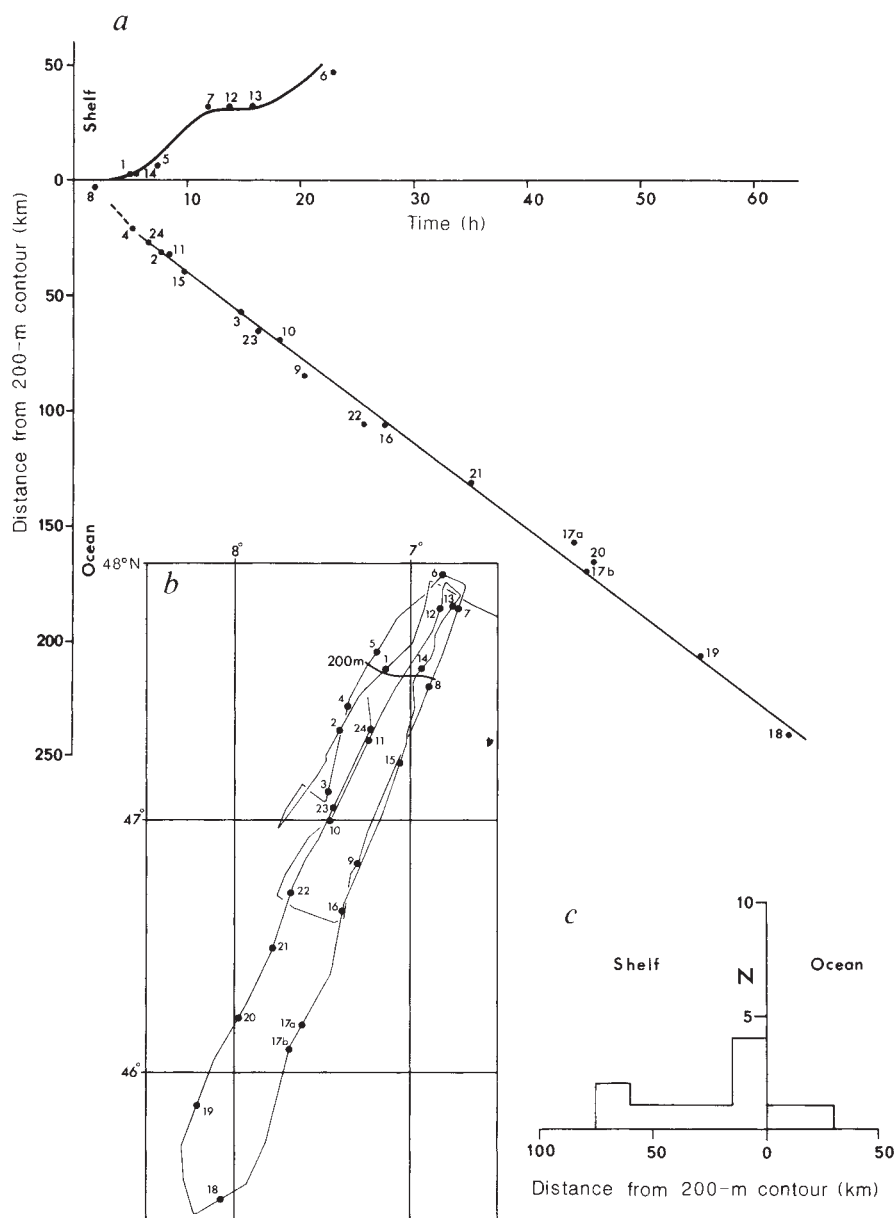


Fig. 4 *a*, Travel-time graphs for all the on-shelf and off-shelf internal tidal troughs identified in the SeaSoar records as the ship progressed along the track shown in *b*. The arbitrary time origin corresponds to high water at Plymouth and distances are measured normally from the shelf break (200-m contour) in kilometres. Individual internal tidal troughs were not followed and so at increasing distance from the shelf break the travel time, t , was determined from $t = T + n\tau$ where T is the time (in h) after the most recent high water (Plymouth) at which a trough was encountered, τ is the semidiurnal tidal period (in h) and n is an integer. Each trough has been numbered in the order in which it was observed and some individual waves were measured subsequently at increasing distances from the shelf break. For example, numbers 11, 16 and 20 represent different encounters with the same tidal trough. Note that the waves are Doppler-shifted, that is, the frequency of encounter is greater when the ship is steaming towards the generating region near the shelf break (compare the observed relative positions of waves 19, 20, 21, 22, 23 and 24 measured when the ship is heading in the direction from which the waves are coming, with the relative positions found for waves 15, 16, 17 and 18 measured as the ship is catching up and overtaking successively older internal tides). Wave 17 appeared to have two troughs and is plotted as 17a and 17b. The on-shelf travel-graph (thick line) represents a steady phase speed of 0.70 m s^{-1} travelling in a medium which is itself oscillating at $\sim 0.70 \text{ m s}^{-1}$ due to the barotropic tidal currents. The off-shelf travel graph is more steady at $\sim 1.03 \text{ m s}^{-1}$ except near the shelf break where the straight line is assumed to have an increased slope (dashed line) to take account of the increased barotropic tidal current measured on the slopes and an assumed approximate mode-1 oscillation. Errors in positioning the internal tidal troughs in space and time are generally $< 3 \text{ km}$ and $< 10 \text{ min}$ respectively so large deviations from the idealized line represent real variability. *c*, Frequency distribution of the number (N) of surface nutrient pulses observed within 15-km intervals from the 200-m contour. The distribution is biased towards the shelf, showing that surface nutrient levels increase at the shelf break faster than over the slopes and oceanic regions as the thermocline is eroded by internal tidal mixing.



lished, and clearly the real situation is more complex, these preliminary results suggest that during the off-shelf streaming phase of the barotropic tide the isotherms in the upper slope region descend and that this depression separates as the on-shelf and oceanward internal tidal troughs propagate from the generating zone^{9,20}. At large spring tides the crest-to-trough height of the internal tide is ~ 50 – 60 m in the generating region and the decay scale (the travel distance over which the amplitude falls to e^{-1} of its initial amplitude) is $\sim 200 \text{ km}$ for the faster ocean-going waves⁸ and $\sim 70 \text{ km}$ for the more slowly propagating shelf internal tide. The more rapid decay rate (double) of the on-shelf tide can be partially attributed to increased dissipation due to the addition of the internal tidal currents to the barotropic tidal currents in the presence of quadratic friction²¹.

Linear theory²² indicates that in a two-layer fluid the phase speed, c , for internal tides with crests that are parallel and horizontal (Sverdrup waves), and whose wavelengths are long compared with the depth of the ocean and whose periods are long enough to be influenced by Earth's rotation, is given by

$$c = (gh_e)^{1/2} \quad (1)$$

where g is the acceleration due to gravity and h_e , the effective depth for the baroclinic mode, is defined as

$$h_e = \frac{\delta\rho}{\rho} \frac{h_1 h_2}{H(1 - f^2/\sigma^2)} \quad \text{with } H = h_1 + h_2 \quad (2)$$

where $\delta\rho$ is the density difference between the upper layer of thickness h_1 , and the lower layer of thickness h_2 , with density ρ . The ratio of the Coriolis parameter, f , at the latitude $\phi = 47^\circ \text{ N}$, to the tidal frequency σ , is $f/\sigma = 0.76$.

The energy contained in these waves propagates at the group velocity, c_g , defined as

$$c_g = \frac{\partial\sigma}{\partial k} = c \left(1 - \frac{f^2}{\sigma^2} \right) = 0.43c \quad (3)$$

where $k = 2\pi/\lambda$ is the wavenumber, indicating that maximum internal tides (spring tides for the internal tide) will tend to occur at a distance l from the shelf break at a time $t \sim 2.3l/c$ after spring tides for the barotropic tides.

Mean values of h_1 and h_2 are estimated in Table 1, where $\Delta\rho$ now refers to the density difference that occurs over an arbitrary

50-m interval centred across the seasonal thermocline. Numerical studies showed that nonlinearity had only a small effect on the phase speed and that if the full density structure was taken into account²³⁻²⁶ then the shelf situation was well represented by the first baroclinic mode, as indicated in Table 1. For the oceanic situation, the density below the thermocline increases slowly with increasing depth, and a two-layer approximation with $\Delta\rho$ estimated as above is not valid. The third and fourth modes give the most realistic phase speeds for the oceanic region. On the upper slopes, in water depths of ~1,500–500 m, the peak-to-trough amplitude of the internal tide increases to ~200–400 m (respectively) at mid-depth and towards the bottom during spring tides. There are some differences of phasing between the upper and lower layers in the deeper regions, and near the bottom in ~500 m depth the currents are intensified at some locations¹⁹, reaching 80–90 cm s⁻¹ at times. Thus, the oceanic situation differs from the shelf situation in that the full density structure results in a generation and conversion of internal tidal energy into higher internal modes and allows rays to penetrate the ocean at an angle given by the internal wave characteristic^{8,19}. The theoretical estimates of the phase speeds for the internal tides are in reasonable agreement with observed values (Fig. 4) of 1.03 m s⁻¹ for the ocean propagating internal tide and 0.7 m s⁻¹ for the on-shelf tide, giving corresponding wavelengths of 46 and 31 km respectively, which supports the interpretation of the data obtained from the sea programme.

During the neap tide period immediately before the observations illustrated in Figs 3 and 4 there was little evidence of any internal tidal activity and the sea-surface levels of inorganic nitrate were uniformly low (<0.1 μ M). During the spring tide period, nutrient pulses (where the surface values increased by >0.1 μ M above background) were detected at the sea surface and their frequency distribution with respect to the 200-m contour is indicated in Fig. 4c. The distribution suggests that most mixing events occur near the 200-m contour but spread well on-shelf (~50 km). Although two nutrient pulses were observed on the slopes the indications are that surface nutrient pulses due to internal waves and tides are rare events in the open ocean at this time of year⁶. The distribution bias towards the shelf, the reduced vertical density structure on the shelf (Table 1) and the more rapid decay of the on-shelf propagating internal tide are consistent with increased mixing on the shelf. Some nutrient pulses are completely mixed to the surface and are detectable on successive passes, others do not persist at the surface and collapse back as the crest of the internal tide (or the crest of a structure of higher wavenumber) passes by. No new nutrient pulses were observed in the region during the next period of neap tides from 20 September to the end of the cruise (26 September).

In the present study, we have established that the thermocline is forced near the shelf break of the Celtic Sea during spring tides. In this process, mixing takes place as the internal tides and higher-wavenumber structures are generated and propagate away from the forcing region. Although a causal relationship between internal tides, nutrient renewal and increases in phytoplankton abundance in the vicinity of the shelf break has been indicated, further studies are required to quantify the nutrient flux and phytoplankton response arising from the internal tidal activity.

We thank EPSHOM and especially Yves Camus for coordinating our participation in *ONDINE*, also John Smithers, Ian Waddington, Bob Longmore and the Research Vessel Services Computing Group for assistance at sea. A.L.N. is supported by an ARE contract with NERC.

Received 5 December 1985; accepted 24 February 1986.

1. Pingree, R. D. *J. mar. biol. Ass. U.K.* **59**, 689–698 (1979).
2. Pingree, R. D. & Mardell, G. T. *Phil. Trans. R. Soc. A* **302**, 663–682 (1981).
3. Pingree, R. D., Mardell, G. T., Holligan, P. M., Griffiths, D. K. & Smithers, J. *Continental Shelf Res.* **1**, 99–116 (1982).
4. Sandstrom, H. & Elliott, J. A. *J. geophys. Res.* **89**, 6415–6426 (1984).
5. Briscoe, M. G. *Nature* **312**, 15 (1984).

6. Holligan, P. M., Pingree, R. D. & Mardell, G. T. *Nature* **314**, 348–350 (1985).
7. Rattray, M. Jr *Tellus* **12**, 54–62 (1960).
8. Baines, P. G. *Deep Sea Res.* **29**, 307–338 (1982).
9. Pingree, R. D., Griffiths, D. K. & Mardell, G. T. *J. mar. biol. Ass. U.K.* **64**, 99–113 (1983).
10. Pingree, R. D. & Mardell, G. T. *Prog. Oceanogr.* **14**, 431–441 (1985).
11. Collins, D. S., Pollard, R. T. & Suchen, P. *Inst. Oceanogr. Sci. Rep.* **148**, 1–75 (1983).
12. Fasham, M. J. R., Pugh, P. R., Griffiths, D. & Wheaton, J. E. *G. Radio electr. Engrn.* **53**, 21–24 (1983).
13. Haury, L. R., Wiebe, P. H., Orr, M. H. & Briscoe, M. G. *J. mar. Res.* **41**, 65–112 (1983).
14. Kerry, N. J., Burt, R. J., Lane, N. M. & Bagg, M. T. *J. phys. Oceanogr.* **14**, 1419–1423 (1984).
15. Sabinin, K. D., Nazarov, A. A. & Serikov, A. N. *Izv. Atmos. Oceanic Phys.* **18**, 317–323 (1982).
16. Madelain, F. & Kerut, E. G. *Oceanol. acta* **1**, 159–168 (1978).
17. Dickson, R. R. & Hughes, D. G. *Oceanol. acta* **4**, 43–46 (1981).
18. Pingree, R. D. in *Remote Sensing of Shelf Sea Hydrodynamics* (ed. Nihoul, J. C. J.) 287–315 (Elsevier, Amsterdam, 1984).
19. Huthnance, J. M. *Prog. Oceanogr.* **10**, 193–226 (1981).
20. Mazé, R. thesis, Univ. Bretagne Occidentale (1983).
21. Hunter, J. R. *Estuar. coast. Shelf Sci.* **3**, 473–479 (1975).
22. Proudman, J. *Dynamical Oceanography* (Methuen, Oxford 1953).
23. Bell, T. H. Jr *Naal Res. Labs, Washington, Rep. No 7294* (1971).
24. James, I. D. *Deep Sea Res.* **29**, 1099–1111 (1982).
25. Pereskokov, A. I. & Schulepov, V. A. *Oceanology* **24**, 565–569 (1984).
26. Pingree, R. D. & Morrison, G. K. *J. phys. Oceanogr.* **3**, 280–285 (1973).

Carbonate recrystallization in basal sediments: evidence for convective fluid flow on a ridge flank

Miriam Kastner, Joris M. Gieskes & Jine-Yu Hu

Scripps Institution of Oceanography, La Jolla, California 92-93, USA

In many Deep Sea Drilling Project (DSDP) drill holes, chalks and/or limestones and even dolomites overlie oceanic basement rocks. These occurrences have often been interpreted as the result of increased calcium carbonate dissolution and reprecipitation as cement and overgrowths (recrystallization) at higher prevailing temperatures, either near an oceanic ridge system or at greater burial depths (burial diagenesis). We report here detailed chemical and isotopic analyses of carbonate sediments recovered from a drill hole on the western flank of the East Pacific Rise during Leg 92 of DSDP. This hole was drilled on 4.6-Myr old oceanic crust and recovered chalks directly overlying oceanic basement at a depth as shallow as 8 m below the sediment/water interface. Oceanic basement was not reached in this hole because of the hardness of deeper chalks, or, more likely, limestones. The chemical and isotopic evidence, detailed below, shows that considerable calcite recrystallization did occur in an aqueous medium similar to normal seawater and at relatively low temperature. This suggests that the observed recrystallization of calcareous ooze to chalk and limestone took place not as a result of burial diagenesis at higher temperatures, but as a result of extensive advective pore water flow through the sediments, which would allow recrystallization of calcium carbonate to take place. This process was, and apparently still is, widespread in the south-east Pacific Ocean.

The sediments of Hole 600C (18°55.70' S, 116°50.45' W) at a water depth of 3,398 m, are at least 11.8-m thick, and consist of mixtures of two components: calcite (CaCO₃) and red to yellow-brown semi-opaque oxyhydroxides¹, typically observed in this area of the East Pacific Rise²⁻⁴. The calcite appears in two forms: as biogenic components and as secondary inorganic precipitates. The relative content of CaCO₃ decreases with depth (Fig. 1). Some basaltic glass and altered basaltic fragments and traces of terrigenous material are also present. At a subbottom depth of 10.25 m a basaltic glass layer occurs which is strongly altered to palagonite and poorly crystalline smectite.

The heat flow at this site is 580 mW m⁻² (almost 14 heat flow units)⁵, with the theoretically expected value being significantly lower⁶ (~245 mW m⁻²). The full spreading rate is ~17 cm yr⁻¹. Interstitial water analyses⁷ indicated only small changes in Ca²⁺ and Mg²⁺ with depth as well as increases in dissolved NO₃⁻. This suggests little influence of recent upward advection in this hole, although in the vicinity of this hole advective flow has been inferred to occur in certain zones of high heat flow⁸.

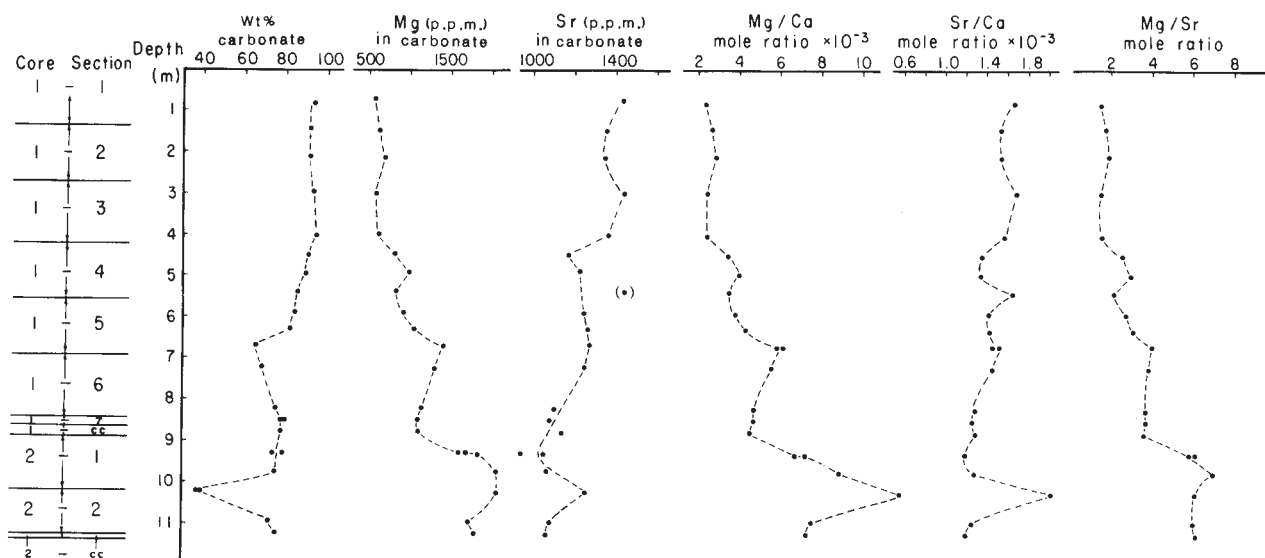


Fig. 1 Weight per cent CaCO_3 , Mg, and Sr concentrations, and Mg/Ca, Sr/Ca, and Mg/Sr mole ratios in carbonate fraction of sediments of DSDP Hole 600C. Age boundaries: 600C-1-1 (145-150), Pleistocene; 600C-1-2 (145-150), early Pleistocene; 600C-1-4 (145-150), early Pliocene (N 20); 600C-2CC, early Pliocene (N 19).

The boundary between Pleistocene and Pliocene is located close to the top of Core 1, Section 4 (Fig. 1), at about 4.5 m sub-bottom depth⁹. A rather abrupt change in sediment colour is observed below this boundary from brownish yellow to dark brown, associated with a decrease with depth in carbonate contents. About 3 m above the base of the Hole calcareous ooze gives way to chalk.

At present, and throughout its history, the sediments of Hole 600C were deposited above the carbonate compensation depth (CCD)^{4,10-13}. Preservation of planktonic foraminifera, however, is only moderate to good in the younger samples and moderate to poor in the Pliocene section¹⁴. The presence of a basal Pliocene chalk layer of at least 3 m thickness in such a thin and young calcareous section cannot be explained by the well documented rates and processes of diagenetic reactions in pelagic calcareous oozes. In this section scanning electron microscopy (SEM) observations revealed extensive calcite dissolution, overgrowth and cement (Fig. 2)¹.

To understand the unexpectedly low degree of preservation of biogenic calcite in the lower two-thirds of Hole 600C, the calcareous fraction was analysed chemically in great detail (Fig. 1). Submarine biogenic calcite recrystallization is generally reflected in chemical, isotopic, and often also in mineralogical changes¹⁵⁻¹⁷.

No important changes in chemical composition are observed in the upper 4-4.5 m, but below this systematic changes do occur, not only in the Sr contents (decreases from ~1,400 to ~1,000 p.p.m. (parts per 10⁶)), but particularly in the Mg contents (increases from ~570 to ~2,000 p.p.m.). Sample 2-2 (4-6 cm) at 10.25 m, that is, volcanic glass layer, is unusual. CaCO_3 content is only ~37%, and both Sr and Mg increase in the recrystallized carbonate, presumably as a result of special conditions associated with reactions involving volcanic glass alteration. In this thin layer, a small fraction of the calcite cement may have originated in conjunction with low-temperature alteration of the volcanic glass¹⁸.

Below 4-4.5 m burial depth changes in Sr/Ca and Mg/Ca ratios appear systematic, as is also demonstrated in Fig. 3, which, in addition, contains information on a sample of a basal limestone recovered from a Lower Miocene sample of DSDP Site 598 of Leg 92. The Sr/Ca ratio in the latter sample (Sr/Ca ~0.63 $\times 10^{-3}$) is close to the equilibrium value at ~2°C in conditions of normal seawater $\text{Sr}^{2+}/\text{Ca}^{2+}$ ratios^{16,19}. We suggest

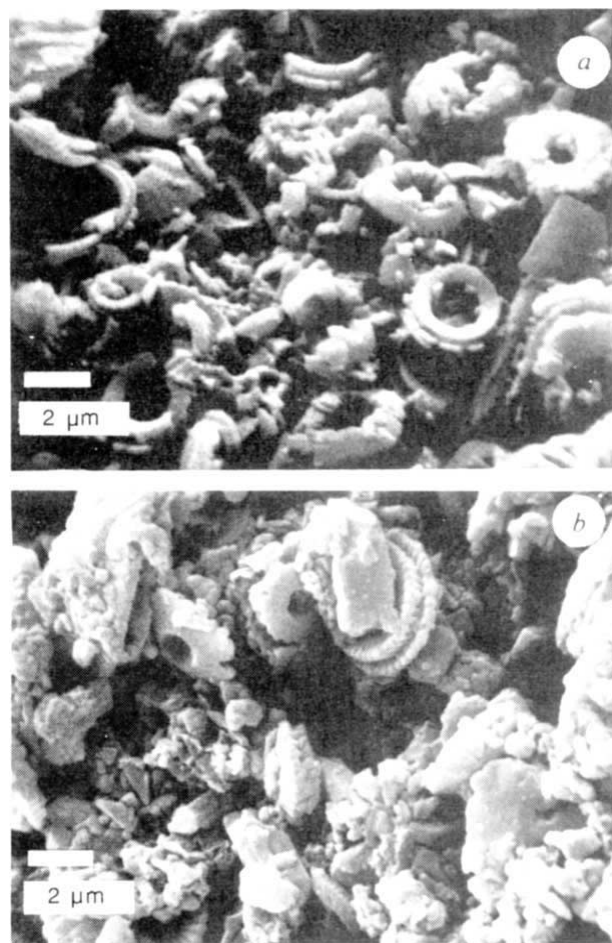


Fig. 2 Scanning electron microscopy photographs of Hole 600C carbonates: *a*, calcareous ooze at 0.8 m depth, and *b*, chalk at 11.2 m depth. Note considerable overgrowth and cementation in chalk but not in the calcareous ooze.

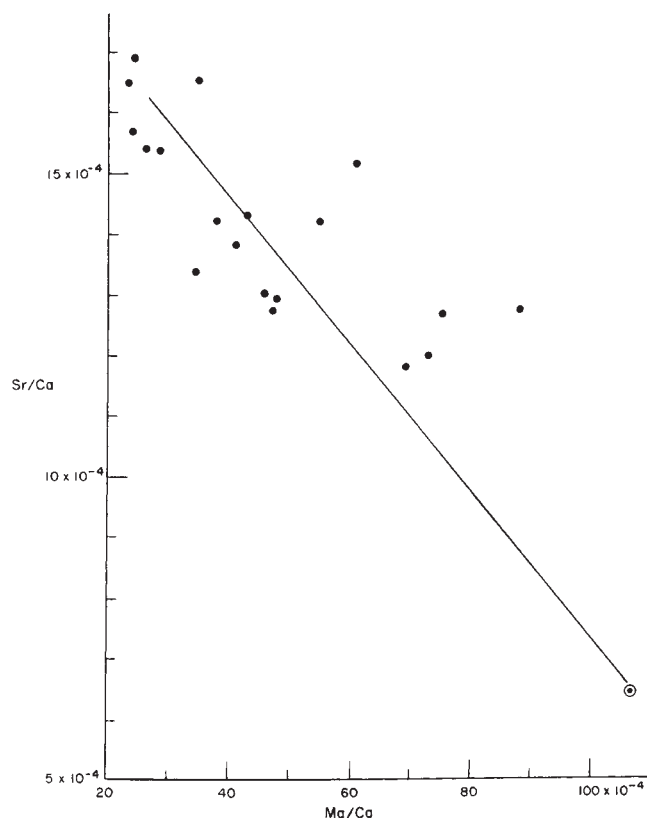


Fig. 3 Correlation between Sr/Ca and Mg/Ca mole ratios in carbonate fraction, DSDP Hole 600C. The circled point is from a basal limestone of nearby DSDP Site 598.

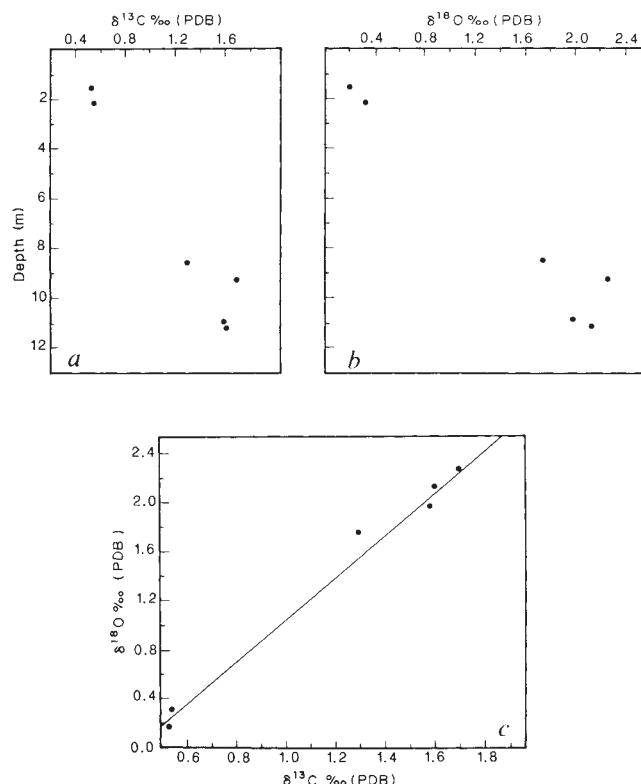


Fig. 4 DSDP Hole 600C: *a*, $\delta^{13}\text{C}$ (PDB) in carbonates versus depth; *b*, $\delta^{18}\text{O}$ (PDB) in carbonates versus depth; *c*, correlation between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ in carbonate fraction.

that the negative slope in Sr/Ca versus Mg/Ca ratios in Fig. 3 indicates a trend of progressive calcite recrystallization.

Values of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ of the bulk carbonate fraction and of the fraction of coccoliths increase significantly with depth (Fig. 4*a,b*). These trends are not expected from the known Pliocene-Quaternary palaeotemperatures²⁰⁻²³. Instead, dissolution of planktonic foraminifera and coccoliths and calcite reprecipitation as overgrowth and cements (recrystallization) at lower than surface water temperatures must be responsible for the observed increases in the $\delta^{18}\text{O}$ values. The $\delta^{18}\text{O}$ value of calcite precipitated from sea water at 2 °C and $\delta^{18}\text{O}_w = -1\text{‰}$ (SMOW), is 1.9‰ (PDB) and at 2 °C and $\delta^{18}\text{O}_w = 0\text{‰}$ is 2.9‰ (ref. 24). As for the carbon isotopes, the $\delta^{13}\text{C}$ of dissolved CO_2 in surface waters is higher than in the deep and bottom waters²⁵. Because of the relatively low ΣCO_2 concentration in seawater ($\sim 2,300 \mu\text{moles kg}^{-1}$ in Pacific Deep Water), in a sea water- CaCO_3 system in which CaCO_3 is being dissolved and reprecipitated, the $\delta^{13}\text{C}$ of the ΣCO_2 will be controlled by the solid carbonate phase. At 25 °C and in conditions of slow rates of precipitation, the carbon of the solid CaCO_3 will be between 1.8 and 2.3‰ heavier than that of dissolved HCO_3^- (refs 26, 27). Up to 60 °C the temperature effect is small, for example, at 20 °C $+(0.035 \pm 0.013)\text{‰ } ^\circ\text{C}^{-1}$ (ref. 26). Thus, during recrystallization at low to moderate temperatures, the $\delta^{13}\text{C}$ value of reprecipitated CaCO_3 will increase. There is a strong positive correlation between the changes in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ (Fig. 4*c*) which again suggests that recrystallization of carbonate is the underlying factor.

On the basis of palaeo-oceanographical and the above documented textural, chemical, and isotopic evidence, we conclude that the processes responsible for the observed extensive calcite dissolution and secondary inorganic reprecipitation and chalk formation in the sediments of Hole 600C, occurred at low temperatures, presumably close to bottom water temperatures, and in solutions similar in composition to bottom sea water.

Indeed, Ca^{2+} and Mg^{2+} concentrations in fluids suggesting upward advective flow in this area are similar to sea water⁸. There is little doubt that the recrystallization occurred at an early age. We propose that observed occurrences of chalks and limestones directly overlying oceanic basement generally form at an early age and that special circumstances are needed to explain their early recrystallization. Furthermore we suggest that the most important cause of this process is advective fluid flow through the sediments on the flanks of midocean ridges. In such conditions, recrystallization reactions are greatly enhanced when compared with diffusion controlled systems in which fluid flow is essentially absent. Thus, our observations in Hole 600C suggest that, at least during the first 1–1.5 Myr of the existence of this sediment section, substantial advective flow must have occurred through the sediments. Increasing degrees of recrystallization with depth presumably imply upward flow from underlying basalts, involving low temperature, relatively unaltered sea water. This cementation will lead to eventual sealing of the basement limiting further advective flow. Much of the oceanic basement in the southeastern Pacific Ocean may be covered by similar limestone or chalk layers, which locally restrict convective fluid flow between sea water and oceanic basement.

We thank all cruise participants of Leg 92 for lively discussions. This research was supported by NSF grants OCE-8300692 (M.K.) and OCE-8218539 (J.M.G.).

Received 4 November 1985; accepted 4 March 1986.

1. Kastner, M., Hu, J.-Y. & Gieskes, J. M. *Init. Rep. DSDP 92* (in the press).
2. Quilty, P. G. *et al. Init. Rep. DSDP 34*, 779–794 (1976).
3. Bass, M. N. *Init. Rep. DSDP 34*, 611–624 (1976).
4. Rea, D. K. & Leinen, M., *Init. Rep. DSDP 92* (in the press).
5. Hobart, M. *Init. Rep. DSDP 92* (in the press).
6. Schlater, J. G., Anderson, R. N. & Bell, M. L. *J. geophys. Res.* **76**, 7888–7915 (1971).
7. Geiskes, J. M. & Boulegue, J. *Init. Rep. DSDP 92* (in the press).
8. Bender, M. L. *et al. Earth planet. Sci. Lett.* **76**, 71–83 (1985/1986).
9. Knuttel, S. & Romine, K. *Init. Rep. DSDP 92* (in the press).
10. Berger, W. H. & Winterer, E. L. *Int. Ass. Sedim. Spec. Publ.* **1**, 11–48 (1974).

11. van Andel, Tj. H., Heath, G. R. & Moore, T. C. *Geol. Soc. Am. Mem.* **143** (1975).
12. Broecker, W. S. & Broecker, S. *Soc. Econ. Paleont. Miner. Spec. Publ.* **20**, 44-57 (1974).
13. Berger, W. H., Adelseck, C. G. & Mayer, L. A. *J. geophys. Res.* **81**, 2617-2627 (1976).
14. Romine, K. *Init. Rep. DSDP* **92** (in the press).
15. Schlanger, S. O. & Douglas, R. G. *Int. Ass. Sedim. Spec. Publ.* **1**, 117-148 (1974).
16. Baker, P. A., Gieskes, J. M. & Elderfield, H. *J. Sedim. Petrol.* **52**, 71-82 (1982).
17. Killingley, J. S. *Nature* **301**, 594-597 (1983).
18. Garrison, R. E., Hein, J. R. & Anderson, T. F. *Sedimentology* **20**, 399-410 (1973).
19. Katz, A., Sass, E., Starinsky, A. & Holland, H. D. *Geochim. cosmochim. Acta* **36**, 481-496 (1972).
20. Douglas, R. G. & Savin, S. M. *Init. Rep. DSDP* **17**, 591-605 (1973).
21. Douglas, R. G. & Savin, S. M. *Init. Rep. DSDP* **32**, 509-520 (1975).
22. Savin, S. & Yeh, H. W. in *The Sea* Vol. 7 (ed. Emiliani, C.) 1521-1554 (Wiley-Interscience) (1981).
23. Berger, W. H., Vincent, E. & Thierstein, H. R. *Soc. Econ. Paleont. Miner. Spec. Publ.* **32**, 489-504 (1981).
24. O'Neil, J. R., Clayton, R. N. & Mayeda, T. K. *J. chem. Phys.* **51**, 5547-5558 (1969).
25. Kroopnick, P., Weiss, R. F. & Craig, H. *Earth planet. Sci. Lett.* **76**, 103-110 (1972).
26. Emrich, K., Ekhardt, D. H. & Vogel, J. C. *Earth planet. Sci. Lett.* **8**, 363-371 (1970).
27. Turner, J. V. *Geochim. cosmochim. Acta* **46**, 1183-1191 (1982).

Mutual diffusion in a miscible polymer blend

R. A. L. Jones, J. Klein* & A. M. Donald

Cavendish Laboratory, University of Cambridge, Madingley Rd, Cambridge CB3 0HE, UK

Self-diffusion in polymer melts has been extensively studied and is reasonably well understood in terms of the reptation model¹⁻⁴; the related phenomenon of mutual diffusion in miscible blends of chemically different polymers has received little attention, despite its practical relevance and implications for physical processes, such as phase separation kinetics. In such blends, attractive interactions between the monomers, when summed over a polymer chain, may lead to large enthalpic driving forces favouring the mixing; this in turn results in a mutual diffusion rate which is rapid compared with the entropically driven self-diffusion⁵, and which is strongly composition-dependent⁶. We have measured mutual diffusion as a function of composition in one such binary blend, polyvinyl chloride (PVC)/polycaprolactone (PCL), and report here that the mutual diffusion coefficient is strongly enhanced in the middle of the composition range. This result is qualitatively, though not quantitatively, in accord with the results of some recent theoretical treatments⁵⁻⁸.

The molecular characteristics of the polymers used are given in Table 1. Diffusion coefficients were determined by measuring the broadening with time of an initially step-like concentration profile. A sharp interface was set up between blends of PVC and PCL differing in concentration by 10%; the sample was then held at a temperature of $91.5 \pm 0.5^\circ\text{C}$ in a vacuum oven for known times (8 h to 7 days). The final distribution of the two polymer species after diffusion-broadening across the interface was then determined using X-ray microanalysis in a scanning electron microscope, in which X rays from inner-shell transitions excited by the electron beam, and characteristic of the element producing them, are counted by a detector. In this case we measured the concentration of chlorine atoms in the PVC across the sample; the spatial resolution of this technique is $3\text{ }\mu\text{m}$, and with careful sample preparation a concentration resolution of 1% is possible. The resulting broadened profile was analysed using the standard Fick's law solutions to the diffusion equation, to yield directly the value of $D(c)$, the mutual diffusion coefficient at the average concentration c of the diffusion couple. Examples of concentration profiles before and after broadening are given in Fig. 1. Previous workers have used a similar technique^{9,10}, but in these earlier studies pure materials were allowed to interdiffuse; this does not give satisfactory information about the concentration dependence of the diffusion coefficient for the polydisperse polymers used in experimental

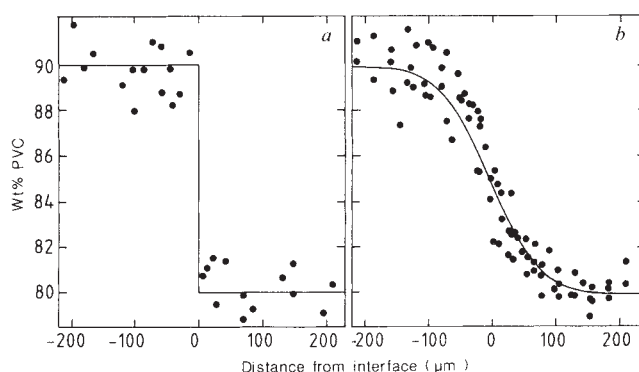


Fig. 1 *a*, Initial concentration profile between two blends (80% PVC/20% PCL, 90% PVC/10% PCL) as revealed by X-ray microanalysis immediately after joining. *b*, Concentration profile after annealing under vacuum for 114.2 h at 91.5°C . The solid curve is an error-function fit corresponding to $D = 7.42 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$, obtained by least-squares analysis. The figure is a composite of five individual profiles measured from different sections of the same diffusion couple. The scatter of each point is ~ 2 -3 times that expected due to counting statistics.

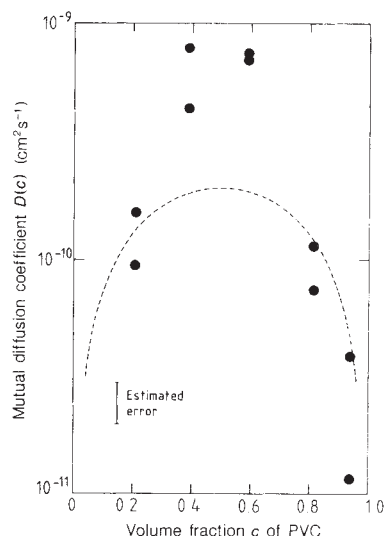


Fig. 2 The measured variation of $D(c)$ with blend composition for the miscible polymer blend PVC/PCL at 91.5°C . The concentration plotted is the mean volume fraction of PVC in the diffusion couple. Each point corresponds to a value of $D(c)$ taken from a composite concentration profile such as that in Fig. 1, made up of several different profiles for the same diffusion time. Points for a given c correspond to runs carried out over times differing by a factor of ~ 3 . The broken line shows the concentration dependence predicted by equation (1), assuming D_0 to be constant and neglecting the difference in degree of polymerization of the polymers. The actual value chosen for D_0 is arbitrary.

Table 1 Relative molecular masses (M_r s) of polymer samples

Polymer	M_w	M_n	M_w/M_n
PVC	83,500	37,400	2.2
PCL	33,000	10,700	3.1

The polymers used in this study were purchased from Aldrich Chemical Co., and were used as supplied. The M_r characterizations, supplied by Aldrich, were obtained by light scattering and gel-permeation chromatography. M_w , weight-averaged M_r ; M_n , number-averaged M_r .

* Permanent address and address for correspondence: Department of Polymer Research, Weizmann Institute, Rehovot 76100, Israel.

work, as the standard methods for obtaining this information from diffusion profiles apply only to strictly binary systems¹¹.

Figure 2 shows the measured variation of $D(c)$ with concentration, based on profiles such as that in Fig. 1. A marked maximum in $D(c)$ towards the centre of the concentration range is apparent.

A prediction for the concentration dependence of $D(c)$ is provided by some recent analyses. As in the case of small molecules, the mutual diffusion coefficient may be written as the product of a mobility term and a thermodynamic term expressing the departure from ideality of mixing¹². The thermodynamic term may be evaluated using the Flory-Huggins theory for the free energy of mixing of a polymer blend; for polymers of identical chain lengths this gives⁶⁻⁸

$$D(c) = D_0[1 - 2\chi Nc(1 - c)] \quad (1)$$

where c is the volume fraction of one polymer, D_0 is a coefficient expressing the mobility of the components (if the two materials have identical mobilities this would correspond to their self-diffusion coefficient), and χ is an interaction parameter (for polymers to be miscible χ must be $<1/2N$). For this system χ has been estimated by melting-point depression, and is expressed per monomer unit of PCL ($\chi = -0.38$)¹³. The presence of the polymerization index N ensures that the second term in equation (1) dominates for all but the smallest concentrations; that is, the diffusion is driven by enthalpy rather than by entropy. It is this that leads to the characteristic observed concentration dependence. For polymers with differing molecular masses the thermodynamic term is slightly different⁸; this introduces a skewing of the curve towards the region rich in the longer polymer.

The measured concentration dependence in the system we have studied is rather stronger than the simple quadratic relation given in equation (1) (compare with broken line in Fig. 2). Quantitative agreement with this result is not to be expected, because of the simplifications in the model used. In particular, no microscopic theory exists for the average mobility term D_0 when the polymers have different mobilities. In this system it is clear that there are large differences between the mobilities of the pure components; this is reflected in the large variation of glass transition temperature with composition¹⁴. Nonetheless, the present study shows directly for the first time that in a miscible polymer blend the mutual diffusion coefficient does indeed vary strongly with blend composition; a maximum is indicated near the middle of the blend composition range, in qualitative agreement with recent theoretical treatments.

We thank Dr F. Brochard, Professor E. J. Kramer and Dr R. C. Ball for useful discussions. R.A.L.J. holds an SERC CASE studentship with ICI PLC.

Received 30 December 1985; accepted 27 February 1986.

1. de Gennes, P. G. *J. chem. Phys.* **55**, 572-579 (1971).
2. Doi, M. & Edwards, S. F. *JCS Faraday Trans. II* **74**, 1789-1801 (1978).
3. Klein, J. *Nature* **271**, 143-145 (1978).
4. Tirrell, M. *Rubber Chem. Technol.* **57**, 523-548 (1984).
5. de Gennes, P. G. *C. r. hebd. Séanc. Acad. Sci., Paris* **292**, 1505-1508 (1981).
6. Brochard, F., Jouffroy, J. & Levinson, P. *Macromolecules* **16**, 1638-1643 (1983).
7. Kramer, E. J., Green, P. F. & Palmstrom, C. J. *Polymer* **25**, 473-479 (1984).
8. Sillescu, H. *Die Makromolekulare Chemie, Rapid Communications* **5**, 599-603 (1984).
9. Price, F. P., Gilmore, P. T., Thomas, E. L. & Laurence, R. L. *J. Polym. Sci. Polym. Symp.* **63**, 33-44 (1978); *Macromolecules* **13**, 880-886 (1980).
10. Chalykh, A. Ye., Sapozhnikova, I. N. & Aliyev, A. D. *Polymer. Sci. USSR* **21**, 1835-1840 (1980).
11. Crank, J. *Mathematics of Diffusion* 2nd edn (Oxford University Press, 1975).
12. Prigogine, I. *Thermodynamics of Irreversible Processes*, 3rd Edn (Wiley, New York, 1967).
13. Woo, E. M., Barlow, J. W. & Paul, D. R. *Polymer* **26**, 763-774 (1985).
14. Koleske, J. V. & Lundberg, R. D. *J. Polym. Sci. Pt A-2* **7**, 795-807 (1969).

Vortex flow visualizations reveal change in upstroke function with flight speed in bats

Jeremy M. V. Rayner, Gareth Jones & Adrian Thomas

Department of Zoology, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

The flapping wings of flying birds and bats generate complex, unsteady air movements. These consist of well-defined vortex structures which are a necessary result of the aerofoil action of the wings, and which transport momentum below and behind the animal in reaction to the lift force which both balances weight and provides thrust¹. Visualization of the vortex patterns enables the aerodynamic function of the flapping wings to be determined²⁻⁶. Here we present the results of experiments with small vesperilionids which allow the first description of the vortex wake of bats. The aerodynamics of flapping flight is similar between birds and bats: thrust is always generated during the downstroke, but wingbeat gait (the cyclic pattern of wing movements) and the mechanical function of the upstroke are determined by wing morphology. We present the first evidence from aerodynamic experiments that in an individual bat or bat species, upstroke function and wingbeat gait also vary with flight speed, with aerodynamic lift being generated during the upstroke at high speeds but not during slow flight. This result confirms that flapping gaits are not species-specific, but are selected according to the mechanical conditions experienced by the animal.

We trained long-eared bats (*Plecotus auritus* L.) and noctules (*Nyctalus noctula* L.) to fly through a cloud of neutrally buoyant helium-filled soap bubbles, and recorded wake airflows by photographing the movements of the bubbles in stereo; illumination by four electronic flash guns fired successively at short intervals (8-12 ms) enabled vortex-induced air speeds to be

determined, since speed is proportional to the distance between bubble images. Discrete vortices are identified by localized high flow velocities and pronounced rotational flows. This method has been applied to various species of birds^{3,4,6}, but has not been used previously with bats. Figures 1-3 show typical photographs from a total of about 250 of the two species, together with sketches, drawn from stereo pairs, of the associated discrete vortex patterns.

Wake photographs should be interpreted in relation to the balance of forces on the flying animal. The lift on the flapping wings acts horizontally as a thrust as well as vertically supporting the weight. If weight support alone were required the wings would not need to be flapped (a fixed gliding wing can generate sufficient lift to equal weight), but flapping of the wings must serve also to balance frictional and vortex drag forces: in gliding flight drag is uncompensated, and the animal has no alternative but to descend or decelerate relative to the air. The wake vortices transport momentum, the mean reaction of which in flapping flight is experienced as a lift force acting on the wings and providing both thrust and weight support. In the downstroke the wings move forwards and downwards relative to the air, and lift has the required forwards and upwards action^{1,5,7,8}. In the upstroke, however, the wings move primarily upwards; at high flight speeds they also travel forwards, the air flow meets the ventral side of the wing as in the downstroke, and if lift is to contribute to weight support it cannot also give thrust, but must retard the animal⁵⁻⁷. In slow flight the wings may move slightly backwards relative to the air; lift forces are likely to be small or absent^{1,7}, but it has been suggested that if in bats the air flow strikes the dorsal surface of the wing, the resulting weak lift may act as both thrust and weight support^{3,7,8}.

Thus, upstroke function is the main mechanism of differentiation between flapping flight gaits in flying vertebrates, and the upstroke may be expected to act in different ways at different flight speeds. Experiments with birds have confirmed that gaits are constrained by wing morphology²⁻⁶: an active, lifting

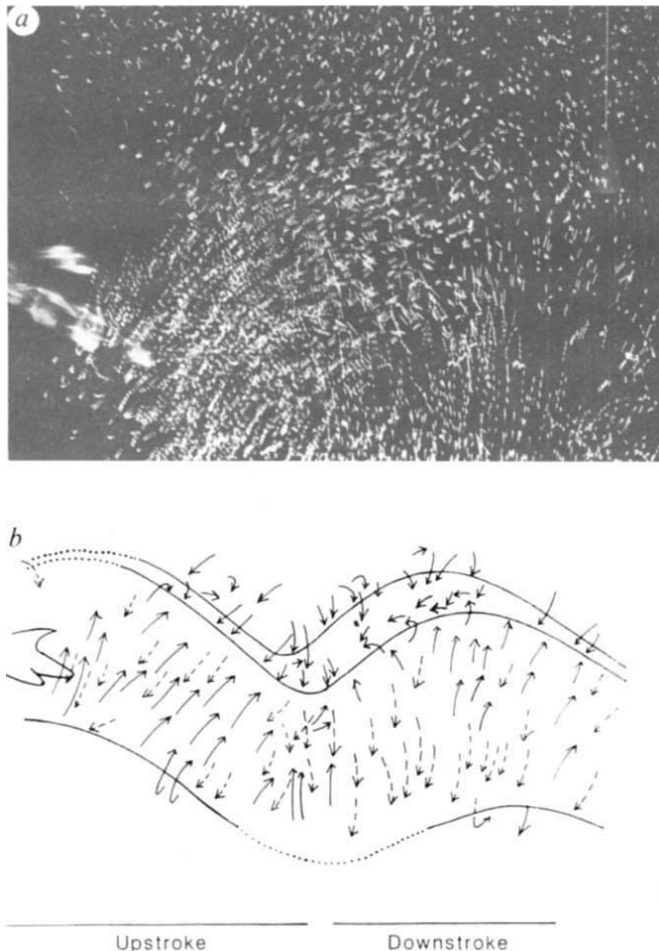


Fig. 1 Flow visualization for the noctule *N. noctula* in steady forward flight at $\sim 7.5 \text{ m s}^{-1}$. *a*, Flow visualization photograph (one half of a stereo pair). *b*, The wake vortices and induced air movements as drawn from stereo pairs. (Bubble direction can be inferred from the photographs as the light intensity is not constant during each flash^{3,4}; broken arrows indicate air movements between the vortex cores.) The wake is formed from a pair of continuous vortex tubes which follow the path of the wings; the spacing of the core centres is $\sim 75\%$ of the distance between the wingtips. The cross-sectional diameter of the cores is greater than that commonly seen in birds²⁻⁴ owing to the relatively small size of the bat. The bat is climbing slightly, but this does not substantially distort wingbeat action compared with level flight. Flash interval 10 ms.

upstroke is more common in long-winged species (typically in fast flight at $\sim 8 \text{ m s}^{-1}$ in the kestrel *Falco tinnunculus*^{3,6}, with above-average aspect ratio and pointed wings), but is absent in small finches², which have shorter, rounded wings. During an active upstroke (that is, with aerofoil action generating a lift force) the effective wingspan must be reduced by flexing or sweeping back the wing to reduce retardation from lift and so ensure a mean positive thrust. Theoretical models predict that in an individual animal the upstroke should be active at higher flight speeds, while in slow flight it should be inactive and the wingtips should be brought close to the body to minimize friction drag and inertial forces^{1,5,7}. The experiments reported here provide the first direct aerodynamic evidence for such changes in upstroke function.

The wakes of flying microbats have much in common with those of birds. When the wings are actively generating lift, the vortex sheets shed from the trailing edges roll rapidly into a pair of discrete vortex lines or tubes behind the wingtips. In fast flight ($7\text{--}9 \text{ m s}^{-1}$ —normal cruising speeds) the noctule's

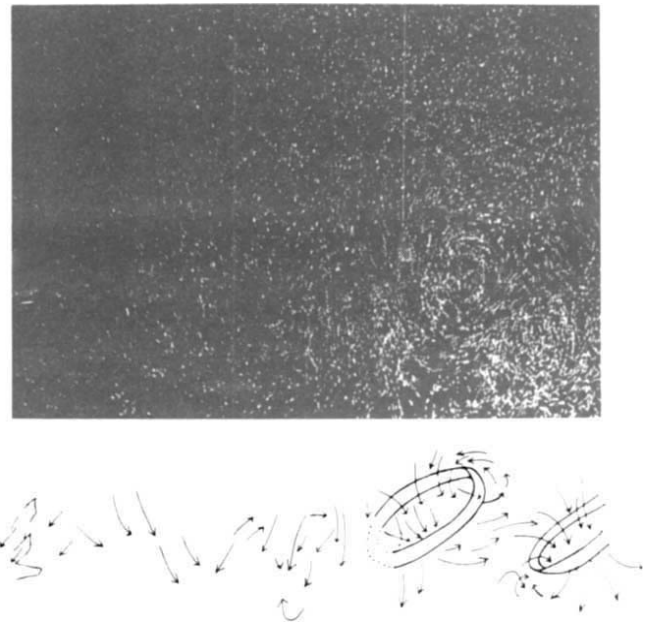


Fig. 2 Flow visualization for noctule in slow flight (3 m s^{-1}), together with the transition to faster flight as the bat accelerates. Two vortex rings are visible to the right; no trailing vorticity is shed during the upstroke, and there is a strong momentum-transferring flux of air through the centres of the rings. These are followed by a weak, continuous tube vortex closer to the bat. Flash interval 10 ms; only three flash guns were used in taking this photograph.

wake is a continuous, undulating vortex pair (Fig. 1); this indicates that the upstroke is active in the way described above, and confirms an earlier prediction⁵ based on wingbeat kinematics. The vortex structure is similar to that seen for the kestrel^{3,6}. The marked absence of transverse vorticity leads to the remarkable conclusion that in both animals the strength (circulation) of the bound wing vortex does not vary during the wingstroke; this conveys certain mechanical advantages^{5,7}. Both noctule and kestrel have wings of above-average aspect ratio, and use similar wingbeat kinematics which are fully consistent with this form of wake⁵. The wingtip is swept back in the bird, while the arm wing is flexed in the bat; but, in both, the wingtip moves towards the body during the upstroke.

At lower speeds (3 m s^{-1}) the noctule wake is rather different (Fig. 2): prominent transverse vortices are shed at the top and bottom of the downstroke, and the upstroke is flexed and inactive. Circulation can no longer remain constant, although it may be nearly so during the downstroke. This 'vortex ring' wake is similar to that of small finches² and slow-flying pigeons^{3,4,6}. Noctules are fast-flying bats which rarely fly sufficiently slowly to show an inactive upstroke, but together our results (as illustrated by Figs 1 and 2) provide the first evidence that upstroke function is variable and is not in general species-specific; moreover, an individual animal changes its gait optimally as flight speed increases. In slow flight ($1\text{--}2 \text{ m s}^{-1}$) in the long-eared bat (which has rather large wings of low aspect ratio) the wake is similar to that of the slow-flying noctule (Fig. 3), with a clear vortex ring corresponding to aerodynamic lift during the downstroke. Our flow visualizations provide no evidence that the upstroke is active in *P. auritus* at any speed, although Norberg concluded from wingbeat photography that in this species at low speeds a weak lift acts as thrust⁸. In slow flight the wingbeat kinematics of both species show considerable flexure and twisting of the wings during the upstroke^{5,7-10}, and any lift generated then is weak. It has been suggested^{5,8,9} that the rapid 'flick' phase at the top of the upstroke in this and similar bats⁸⁻¹⁰ may

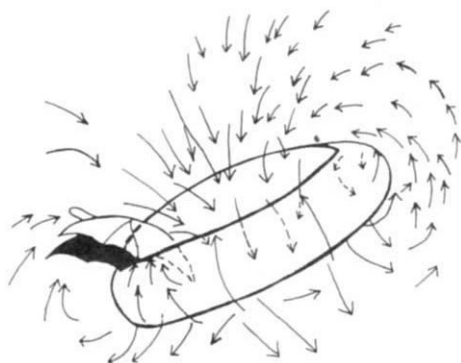
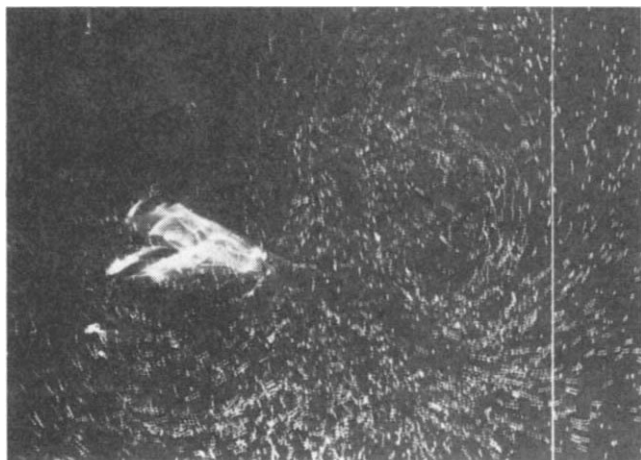


Fig. 3 Flow visualization for long-eared bat *P. auritus* in slow flight (1.5 m s^{-1}). Intense rotational flows induced by a vortex ring generated during the downstroke are visible; at the time of the photograph the wings were at the bottom of the downstroke, and ring generation was still in progress. No vorticity had been shed during the preceeding upstroke, but a similar ring was generated by the preceeding downstroke. The wake in this species is similar to that recorded previously in finches² and pigeons^{3,4}. Flash interval 8 ms. The matching stereo pair of this photograph is shown on the front cover.

generate thrust; this would entail generation of a concentrated transverse vortex at the start of the flick, approximately midway through the upstroke. We found no trace of such vortices in any of our photographs, although as any lift would necessarily be small, they might have been obscured by the more dominant forces and vortices associated with the downstroke. Although unsupported by the present experimental evidence, we cannot exclude the possibility of weak upstroke lift in at least some bat species.

Despite obvious morphological differences, the flapping-wing aerodynamics of birds and bats show considerable similarities. In both groups gaits are defined by upstroke function, which has been shown to vary with wing morphology. Gait changes with flight speed similar to those reported here in bats have been inferred in birds from analysis of wingbeat kinematics^{1,5}; similarly, in bats these changes are likely to be reflected in generation of upstroke lift and the structure of the wake vortices.

We thank H. D. J. N. Aldridge, M. A. Rezende and G. R. Spedding for advice, and U. M. Norberg for comments on the manuscript. This work was supported by grants from the Royal Society, NERC and SERC.

Received 6 February; accepted 20 March 1986.

1. Rayner, J. M. V. *J. exp. Biol.* **80**, 17–54 (1979).
2. Kokshaysky, N. V. *Nature* **279**, 146–148 (1979).
3. Spedding, G. R. Thesis, Univ. Bristol (1982).

4. Spedding, G. R., Rayner, J. M. V. & Pennycuik, C. J. *J. exp. Biol.* **111**, 81–102 (1984).
5. Rayner, J. M. V. in *Biona Report 5* (ed. Nachtigall, W.) 27–74 (Gustav Fischer, Stuttgart, in the press).
6. Rayner, J. M. V. in *Konstruktionsprinzipien lebender und ausgestorbener Reptilien. Konzepte SFB 230 4* (eds Riess, J. & Frey, E.) 89–118 (University of Tübingen, 1985).
7. Rayner, J. M. V. in *Recent Advances in the Study of Bats* (eds Fenton, M. B., Racey, P. A. & Rayner, J. M. V.) (Cambridge University Press, in the press).
8. Norberg, U. M. *J. exp. Biol.* **65**, 179–212 (1976).
9. Eisentraut, M. *Z. wiss. Zool.* **148**, 159–188 (1936).
10. Aldridge, H. D. J. N. Thesis, Univ. Bristol (1985).

Genetics and evolution of female choice

M. E. N. Majerus, P. O'Donald, P. W. E. Kearns & H. Ireland

Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK

Fisher's theory of the evolution of sexual selection by female choice depends crucially on the premise that the preference of females to mate with males of a specific genotype is itself genetic¹. In the polymorphic ladybird *Adalia bipunctata*, a genetically determined, non-assortative female preference for melanic males has been demonstrated². Models of the evolution of female choice show that the rate of evolution and the final outcome of this selection depend critically on the exact mode of inheritance of both the preferred character and the preference³. Here we describe experiments in which the level of preference has been raised by artificial selection over 14 generations. Resultant estimates of heritability are consistent with models in which one or a small number of genes control the preference. Analysis of the levels of preference in isofemale lines derived from the high preference stock shows a quadrimodal distribution in preference. These results are entirely consistent with the deduction that a single dominant gene controls female choice. We conclude that this complex behavioural strategy is based on simple genetics. Furthermore, the demonstration of a simple genetic basis to female mating preferences in *A. bipunctata* implies that sexual selection by female choice is not only important in the evolution of male sexual adornments in sexually dimorphic species, but may also maintain polymorphisms that are not sex-limited.

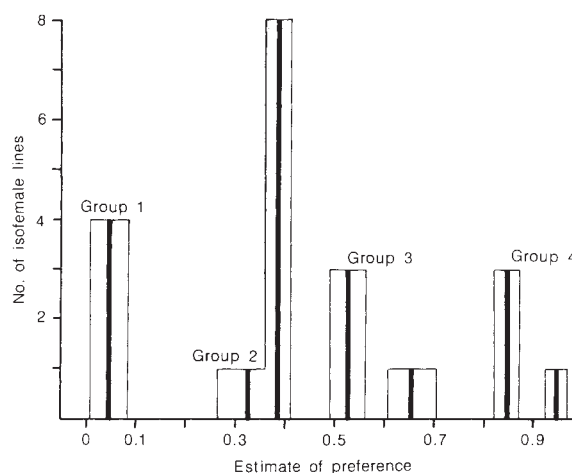


Fig. 1 Isofemale lines shown divided into four distinct groups. The solid black bar shows the position of the estimated preference in the given number of lines, open bars, the standard error. In groups 2, 3 and 4, the preference is also shown for the line most divergent from the others in the group. When groups are considered separately, each of groups 1, 2 and 3 is found to be statistically homogeneous. Group 4 is heterogeneous, with line Z35 having a significantly higher preference. In later population cage experiments, however, Z35 did not show a higher preference. The total χ^2 for heterogeneity within all groups is not significant (Table 2).

Table 1 Matings of males in experiments to select for increased preference in favour of melanic Q males

No. of matings and estimates of preference		
Generation	No. of matings Q♂♂ T♂♂	Estimates of preference $\hat{\gamma} \pm \sqrt{\text{var}(\hat{\gamma})}$
1	76 98	0.195 ± 0.054
2	47 39	0.352 ± 0.077
3	67 53	0.369 ± 0.065
4	73 48	0.433 ± 0.064
5	132 94	0.406 ± 0.047
6	129 87	0.425 ± 0.048
7	117 63	0.500 ± 0.051
8	125 48	0.604 ± 0.049
9	125 41	0.647 ± 0.048
10	96 55	0.480 ± 0.056
11	118 49	0.581 ± 0.050
12	153 77	0.522 ± 0.044
13	130 61	0.544 ± 0.048
14	118 41	0.632 ± 0.050

Analysis of χ^2 of variation in preference

Component of variation	χ^2	Degrees of freedom	P
Regression on generations	47.2017	1	—
Residual	22.8445	12	0.0291
Total for 2 × 14 table	70.0462	13	—

Q represents the four-spot melanic *A. b. quadrimaculata* and T the two-spot non-melanic *A. b. typica*. The numbers of matings are the combined numbers in two replicate lines. $\hat{\gamma}$ is the estimated proportion of preferential matings for Q males, shown with the standard error¹⁵. At generations 5 and 10, the replicates were allowed to mate together without artificial selection. Two new lines were then formed from the unselected generations 5 and 10 to give selected generations 5 and 10 for each of the new replicates. The data shown are those for the selected generations 5 and 10 only. The drop in preference at generations 5 and 10 suggests that there may have been some natural selection against the high preference. The latter drop is the main source of the heterogeneity that gives rise to the significant residual χ^2 .

In ladybirds, sexual selection takes place because females exercise a genetically determined preference for melanic males. In a natural population of the two-spot ladybird, *A. bipunctata*, at Keele, UK, melanic males gain a mating advantage over non-melanic males. About 20% of matings take place preferentially with melanic males¹. Majerus, O'Donald and Weir² selected increased female preference, raising the proportion of preferential matings from 20% to more than 50% in four generations. Reciprocal matings between males and females from the selected, high preference line and the unselected, original Keele stock showed that female preference, and not male competitiveness, had indeed been selected. The preference was genetic, in accordance with the premise of Fisher's theory.

We have now repeated this experiment, but have selected for both increased and reduced preference, producing high and low preference lines. The experimental protocol follows that of our previous selection experiment². The results of selecting for increased preference are shown in Table 1. At generation 10, single pairs were selected at random from the high line. Each batch of eggs was laid by each female in a separate Petri dish and the larvae thus reared separately to avoid larval cannibalism. In this way, large numbers of progeny, up to 600 from some females, have been produced for measurement of the female preference within each 'isofemale line'. In initial, formal mating tests, 10 males and 10 females each consisting of 5 *A. bipunctata quadrimaculata* melanics (Q) or 5 *A. bipunctata sexpustulata* melanics (S) and 5 *A. bipunctata typica* non-melanics (T) were placed in containers and allowed to mate. At 30-min intervals, mating pairs were removed and replaced with individuals of

Table 2 Initial formal mating tests on mating preferences in isofemale lines

Frequencies of matings with males		Mated males		Total males		Estimate of preference
Group	Line	Q	T	Q	T	
1	Z4	51	48	210	210	0.033 ± 0.088
	Y4	218	204	815	815	0.038 ± 0.042
	Z33	49	45	220	220	0.046 ± 0.091
	Y12	57	52	245	245	0.050 ± 0.084
	Totals	375	349	1,490	1,490	0.041 ± 0.032
2	Y18	134	73	395	395	0.324 ± 0.055
	Y14	149	74	435	435	0.367 ± 0.052
	Z36	141	69	400	400	0.375 ± 0.053
	Y13	223	107	700	700	0.380 ± 0.043
	Z17	212	102	620	620	0.381 ± 0.044
	Z34	141	66	420	420	0.393 ± 0.054
	Z20	68	33	125	125	0.398 ± 0.067
	Y15	154	70	395	395	0.411 ± 0.049
3	Z32	74	32	200	200	0.430 ± 0.072
	Totals	1,296	626	3,690	3,690	0.381 ± 0.018
4	Z13	218	79	590	590	0.501 ± 0.042
	Y9	150	50	360	360	0.537 ± 0.048
	Z16	80	26	200	200	0.544 ± 0.067
	Y3	153	36	335	335	0.654 ± 0.044
4	Totals	601	191	1,485	1,485	0.553 ± 0.024
	Y19	272	34	575	575	0.801 ± 0.028
	Z37	91	6	240	240	0.887 ± 0.040
	Z40	102	5	200	200	0.918 ± 0.031
	Z35	203	7	385	385	0.942 ± 0.018
4	Totals	668	52	1,400	1,400	0.872 ± 0.015

Analysis of χ^2

Component of variation	χ^2	Degrees of freedom	P
Between groups	362.39	3	—
Within groups	24.69	17	0.102
Within replicates of lines	20.35	19	0.374
Total	407.43	39	—

In this analysis of χ^2 , the data for replicates of tests on individual lines have not been included. Totals over replicates within lines are shown. The analysis of χ^2 is based on the null hypothesis of random mating of male phenotypes: it is an analysis of the values of χ^2 for interaction in the $2 \times 2 \times N$ table of numbers of mated and unmated phenotypes of males in tests on the isofemale lines. Note that the same males and females may be used several times in successive series of tests.

similar phenotypes. In each 30-min interval, therefore, matings took place without replacement. The results are shown in Table 2. Since as many as five or six matings might take place in each interval, the sampling without replacement creates a very difficult problem for the statistical estimation of the preferences. Nevertheless, a general test of significance based on the null hypothesis of no preference shows wide variation between the isofemale lines. Female progeny from some lines mate almost solely with melanics, while those from others mate at random or show intermediate levels of preference (between lines $\chi^2_{20} = 387.08$; $P \approx 0$). No significant differences were found when Keele males or males from other lines were used instead of males from the same line as the females ($\chi^2_4 = 6.01$; $P = 0.198$). The variation in preference is thus produced entirely by genetic differences contributing to the female choice. Lines showing high, intermediate and low levels of preference were chosen for further genetic analysis. We have now tested the preference in some of these lines using large numbers in population cages, with replacement of males as in our original experiments². Table 3

gives the results obtained so far.

The distribution of levels of preference in the isofemale lines suggests that female preference for melanic males may be largely under the control of one or a very few major genetic loci. The isofemale lines can be divided into four groups according to their preference (see Table 2, Fig. 1). Group 1 females show no significant preference, group 2 females have levels of preference between 0.32 and 0.43, group 3 females have levels of preference of 0.50–0.65, and group 4 females have the highest levels of preferences, between 0.80 and 0.94. There is overall homogeneity within the groups ($\chi^2_{17} = 24.69$; $P = 0.102$), and highly significant differences between groups ($\chi^2_3 = 362.39$; $P = 1.8 \times 10^{-9}$).

The population cage experiments on progeny from the isofemale lines corroborate this apparent discontinuity in the preference distribution (Table 3). Matings were observed in cages in which melanic males and non-melanic males were maintained at the ratio 3Q:7T as in previous experiments^{2,4}. If matings take place with males at this ratio, no preference has been exercised and mating has been random. An excess of Q males in mating indicates preference. A simple model yields an estimate of the proportion of females mating preferentially

$$\hat{\gamma} = \frac{n_Q(1-v) - n_T v}{(1-v)(n_Q + n_T)}$$

where v is the frequency of Q males ($v = 0.3$ in these experiments), and n_Q and n_T are respectively the numbers of Q and T males observed to mate. Again, some lines show high levels of preference (lines Y19, Z28, Z35, Z37 and Z40), while others show no preference at all (lines Y12, Z4, Z33) or intermediate levels of preference (lines Y9, Y14, Z13, Z34).

The pattern of preference levels can be easily explained by a single dominant gene. In the population at Keele, females would either carry the preference gene and hence show the high level of preference at about $\gamma = 0.85$ or lack the gene and show no preference at all. If about one-quarter of the females have the preference, this would give the observed level of 20% preferential mating in the unselected Keele population. After selection has raised the preference to around 50% as in our high line, the preference gene would then have a frequency of $P = 0.36$. At this gene frequency, 24% of matings should only produce females with the preference ($\gamma = 0.85$), 21% should produce females of which three-quarters have the preference ($\gamma = 0.64$) and 38% of matings should produce females of which half have the preference ($\gamma = 0.43$). Females from the rest of the matings should all mate without preference, for none of them would carry the preference gene. The isofemale lines follow this pattern closely, as Fig. 1 shows, with groups 1–4 represented by 4, 9, 4 and 4 isofemale lines respectively ($\chi^2_3 = 0.453$, $P = 0.93$ for test of goodness-of-fit).

We have thus shown that female ladybirds are highly polymorphic in their preference for melanic males. This polymorphism is genetic. Single-gene models, such as those of O'Donald³ and Kirkpatrick⁵, may therefore be sufficient to explain the evolution of female choice as a mating strategy.

In the ladybirds, natural and sexual selection apparently act to maintain a stable polymorphism of melanic and non-melanic phenotypes⁶. However, unlike the most obvious and widely quoted examples of evolution by sexual selection—such as the breeding plumage of male peacocks, pheasants, ruffs and birds of paradise—the sexes of the ladybirds are alike: they both show the same polymorphism of melanic and non-melanic phenotypes.

The theory of sexual selection may thus apply more widely than has generally been supposed. It may explain not only the evolution of highly developed display characters in males but also the evolutionary stability of simple genetic polymorphisms in which no sexual dimorphism is present. Sexual selection has been implicated in the maintenance of other non-sex-limited polymorphisms, in scarlet tiger moths⁷, snow geese^{8,9}, arctic

Table 3 Female preferences from population cage experiments on isofemale lines

Tests on lines producing both Q and T males

Lines used in tests		No. of matings		Estimate of preference $\hat{\gamma} \pm \sqrt{\text{var}(\hat{\gamma})}$
Females	Males	Q/S	T	
Y9(Q, T)	Y9(Q, T)	67	39	0.474 $\pm$ 0.067
Y14(Q, T)	Y14(Q, T)	58	52	0.325 $\pm$ 0.068
Z13(Q, T)	Z13(Q, T)	75	40	0.503 $\pm$ 0.063
Z28(S, T)	Z28(S, T)	100	6	0.919 $\pm$ 0.032
Z34(Q, T)	Z34(Q, T)	52	57	0.253 $\pm$ 0.068
Z37(Q, T)	Z37(Q, T)	89	13	0.818 $\pm$ 0.047
Z37(Q, T)	Keele(Q, T)	92	8	0.886 $\pm$ 0.039
Z37(T)	Keele(Q, T)	86	16	0.776 $\pm$ 0.051

Analysis of χ^2

Component of variation	χ^2	Degrees of freedom	P
Between lines	124.8885	5	—
Between males in line Z37	1.5329	2	0.465
Total	126.4214	7	—

Tests on lines producing only one phenotype

Lines used in tests		No. of matings		Estimates of preference $\hat{\gamma} \pm \sqrt{\text{var}(\hat{\gamma})}$
Females	Males	Q	T	
Y12(T)	Keele(Q, T)	26	73	0.053 $\pm$ 0.063
Y19(T)	Keele(Q, T)	86	19	0.741 $\pm$ 0.054
Z4(T)	Keele(Q, T)	33	66	0.048 $\pm$ 0.068
Z33(Q)	Keele(Q, T)	38	59	0.131 $\pm$ 0.071
Z33(Q)	Keele(Q, T)	34	69	0.043 $\pm$ 0.066
Z35(T)	Keele(Q, T)	94	8	0.888 $\pm$ 0.038
Z40(T)	Keele(Q, T)	89	9	0.869 $\pm$ 0.042

Analysis of χ^2

Component of variation	χ^2	Degrees of freedom	P
Between lines	220.3729	5	—
Within line Z33	0.7744	1	0.379
Total	221.1473	6	—

S represents the six-spot melanic *A. b. sexpustulata*, which occurs in line Z28. 'Keele' means the original unselected stock from Keele, U.K. Letters in parentheses show the phenotypes used in the test.

skuas¹⁰, and even in the polymorphism of electrophoretic allozymes in the milkweed beetle *Tetraopes tetraophthalmus*¹¹. However, the genetic basis of the mating preference has not been demonstrated in any of these cases. That female mating preference can be controlled by a simple genetic system could allow for the appearance and rapid evolution of a frequency-dependent mechanism that would maintain such polymorphisms.

In the wider context of the sociobiological interpretation of animal behaviour, we have shown that a single gene may determine behaviour as complex as the alternative mating choices that an animal might make. This has been a point of great controversy. Models in sociobiology have been based on the assumption that alternative strategies—for example whether to be a 'hawk' or 'dove' in conflicts—are genetic polymorphisms^{12,13}. This simple genetic model of complex behaviour has been challenged by opponents of sociobiology¹⁴, but female choice behaviour in *A. bipunctata* is certainly both genetic and polymorphic.

This work was funded by a SERC research grant to P.O'D. and was carried out in a laboratory built in 1984 with a grant from the Wolfson Foundation for research in population genetics.

Received 20 December 1985; accepted 12 March 1986.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Majerus, M. E. N., O'Donald, P. & Weir, J. *Nature* **300**, 521-523 (1982).
3. O'Donald, P. *Heredity* **22**, 499-578 (1967); *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
4. Majerus, M. E. N., O'Donald, P. & Weir, J. *Heredity* **49**, 37-49 (1982).
5. Kirkpatrick, M. *Evolution* **36**, 1-12 (1982).
6. O'Donald, P., Derrick, M., Majerus, M. E. N. & Weir, J. *Heredity* **52**, 43-61 (1984).
7. Sheppard, P. M. *Heredity* **6**, 239-241 (1952).
8. Cooke, F. & Cooch, F. G. *Evolution* **22**, 289-300 (1968).
9. Cooke, F., Mirsky, P. J. & Seiger, M. B. *Can. J. Zool.* **50**, 529-536 (1972).
10. O'Donald, P. *The Arctic Skua* (Cambridge University Press, 1983).
11. Eanes, W. F., Gaffney, P. M., Koehn, R. K. & Simon, C. M. *Lect. Notes Biomath.* **19**, 49-64 (1977).
12. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).
13. Maynard-Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, 1982).
14. Lewontin, R. C. *Nature* **266**, 283-284 (1977).
15. O'Donald, P. & Majerus, M. E. N. *Heredity* **55**, 401-412 (1985).

A benzodiazepine used in the treatment of insomnia phase-shifts the mammalian circadian clock

Fred W. Turek & Susan Losee-Olson

Department of Neurobiology and Physiology,
Northwestern University, Evanston, Illinois 60201, USA

Between 5 and 20% of the adult population in Western countries suffer from insufficient and/or unsatisfying sleep^{1,2}, often associated with certain psychiatric disorders or with certain types of professional activities (for example, shift workers) and travel schedules (for example, jet lag)^{1,3,4}. The benzodiazepines are at present the drug treatment of choice for the management of anxiety and stress-related conditions as well as insomnia⁵. Benzodiazepines are thought to act by potentiating the action of the neurotransmitter γ -aminobutyric acid (GABA), a widely distributed transmitter in the central nervous system⁶. The circadian system has a key role in the regulation of the sleep-wake cycle⁷⁻¹², and at least some forms of insomnia may be the result of a disorder of the circadian sleep-wake rhythm^{8,9}. Similarly, at least some forms of depression may also involve disruption of normal circadian rhythmicity^{13,14}. A central pacemaker for the generation of many circadian rhythms in mammals, including the sleep-wake cycle, appears to be located in the suprachiasmatic nucleus¹⁵, and recent research indicates that both cell bodies and axons containing GABA are present within the bilaterally paired suprachiasmatic nuclei^{16,17}. These findings raise the possibility that the benzodiazepines, commonly prescribed for sleep and mental disorders, may have an effect on the central circadian pacemaker. Here we report that the acute administration of triazolam, a short-acting benzodiazepine¹⁸ commonly prescribed for the treatment of insomnia, induces a phase-shift in the circadian rhythm of locomotor activity in golden hamsters. This suggests a role for GABA-containing neurones in the mammalian circadian system.

A useful approach in determining whether a specific pharmacological agent can influence a circadian clock is to follow an easily measured pacemaker property, such as the phase of a rhythm driven by the clock, during acute administration of the drug¹⁹. We have examined the potential phase-shifting effects of triazolam on the circadian rhythm of wheel-running behaviour in golden (Syrian) hamsters. Nine-week-old outbred adult male golden hamsters (*Mesocricetus auratus*, Lakeview Hamster Colony, Newfield, New Jersey) were initially housed in groups and placed under a light/dark (LD) regime of 14:10 (14 h of light per 24-h day) for 1-3 weeks. The animals were then moved to individual cages equipped with a running wheel to allow for the continuous recording of locomotor activity²⁰. The animals were transferred 2 to 3 weeks later to either constant light (LL; $N = 40$) or constant darkness (DD; $N = 40$) and were kept under these constant environmental conditions for at least 2 weeks

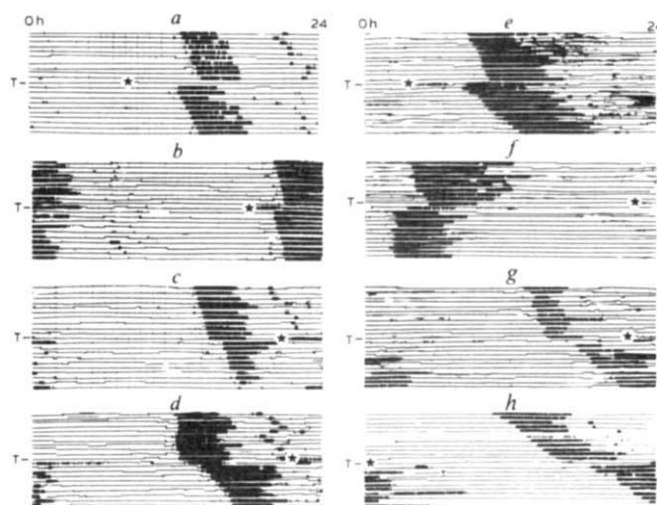


Fig. 1 Representative sections from the activity records of eight hamsters for a 10-12-day period before and after they were injected with 2.5 mg of triazolam at various circadian times. Each line represents a single 24-h period and successive days are plotted from top to bottom in each record. The four animals depicted on the left (a-d) were exposed to constant darkness while the four animals shown on the right (e-h) were exposed to constant light throughout the study. The day of the triazolam injection is noted by the letter 'T' to the left of each record, and the exact time of the injection on that day is indicated by the star. Using the daily onsets of activity for the 5 days preceding the injection as phase reference points for the pre-injection rhythm, a time of activity onset on the succeeding day was graphically projected. In addition, the daily onsets of activity for 2-7 days after the injection (once a new steady state was achieved) were used as phase reference points for the post-injection rhythm, and a time of activity onset for the first circadian day following the injection was projected backwards. The subtraction of the estimated phase of the rhythm after the treatment from the projected phase of the rhythm before treatment yields the magnitude of the phase-shift induced by the injection of vehicle alone or triazolam. Thus, a negative value indicates a delay in the onset of activity; a positive value indicates a phase advance. Using the onset of activity as representing circadian time (CT) 12, the CT of each injection and the resulting phase shift ($\Delta\phi$) in minutes were calculated to be: a, CT=6, $\Delta\phi = +150$; b, CT=9, $\Delta\phi = +90$; c, CT=18, $\Delta\phi = -60$; d, CT=21, $\Delta\phi = -85$; e, CT=6, $\Delta\phi = +110$; f, CT=6, $\Delta\phi = +180$; g, CT=18, $\Delta\phi = -120$; h, CT=21, $\Delta\phi = -100$. On 32 different occasions, vehicle alone was injected at various circadian times. No clear phase-shift (>10 min) was observed in the activity rhythm following any of these injections.

before receiving an intraperitoneal injection of vehicle or 2.5 mg of triazolam dissolved in 0.5 ml of dimethyl sulphoxide at various circadian times. All injections were timed to occur 0, 3, 6, 9, 12, 15, 18 or 21 h after the onset of activity, with onset of activity being designated as circadian time (CT) 12. Thus CT 6 and 21 represent phase points occurring respectively 6 h before and 9 h after the onset of locomotor activity. The hamsters were injected with vehicle or triazolam at least once, and up to four times, with an interval of at least 2 weeks between consecutive injections. The phase-shifting effects of triazolam on the circadian rhythm of locomotor activity were assessed by comparing the onsets of activity before and after drug treatment. The precise nature of the onset of wheel-running activity in the golden hamster makes it easy to detect phase-shifts as small as 15 min (Fig. 1).

While injections of vehicle alone had no effect on the rhythm of locomotor activity, injections of 2.5 mg of triazolam induced phase-shifts in the activity rhythm that were dependent on the circadian time of administration in animals exposed to either

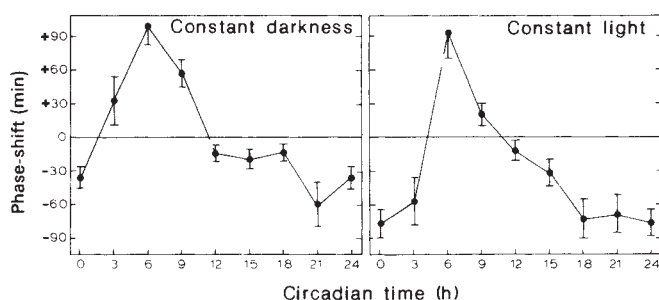


Fig. 2 Mean ($\pm$ s.e.m.) phase-shifting effect of a systemic injection of triazolam at various circadian times on the free-running rhythm of locomotor activity in hamsters exposed to constant darkness (left, 9 animals per time point) or constant light (right, 7–9 animals per time point). A point above the solid line indicates an advance in the onset of activity, a point below the line indicates a delay in activity onset. The onset of locomotor activity represents CT 12. The values presented for circadian times 0 and 24 represent the same data.

DD or LL (Fig. 1). Phase-shifts induced by triazolam were usually completed within one or two cycles. Injection of as little as 0.1 mg of triazolam has also been found to phase-shift the circadian rhythm of activity (F.W.T., unpublished results). A plot of the phase-shifts induced by triazolam as a function of the circadian time of administration revealed that the general shapes of the 'phase response curves' in DD and LL were similar, with triazolam injections at CT 6 and 9 inducing phase advances while injections between CT 12 and 24 induced phase delays (Fig. 2). The major difference between the phase response curves is that while the mean response to an injection of triazolam at CT 3 was an advance of the activity rhythm for animals exposed to DD, a delay in the activity rhythm was observed in the animals exposed to LL that were injected with triazolam at this circadian time.

The phase response curves generated by injections of triazolam are radically different from those generated by light pulses (on a background of constant darkness) in hamsters and in other mammalian species^{20,21}. Light pulses, as well as carbachol, a cholinergic agonist that mimics the phase-shifting and entraining effects of light pulses, induce phase delays early in the subjective night (for example CT 12–14) and phase advances late in the subjective night (for example CT 18–22)¹⁹. The region CT 3–12 is referred to as a 'dead zone' because light does not induce any phase-shift at these circadian times²¹. In contrast, triazolam induces clear phase advances during the dead zone for light and induces phase delays at a time when light induces phase advances. These results, together with the fact that the phase response curves for triazolam are similar under either LL or DD conditions, indicate that the phase-shifting effects of triazolam do not involve an activation of light input pathways to the biological clock.

Because the effects of triazolam are probably mediated by the GABAergic system⁶, the present results suggest a role for GABA-containing neurones in an input pathway to a central biological clock regulating the circadian rhythm of locomotor activity (and presumably many other circadian rhythms) and/or in the clock system itself. The finding that GABA neurones are located within a central circadian pacemaker in rodents—the suprachiasmatic nucleus^{16,17}—raises the possibility that the phase-shifting effects of triazolam occur at the level of the suprachiasmatic nucleus itself. Further support for a role of GABA in the circadian organization of mammals is the observation that bicuculline, a selective antagonist of GABA, can block the phase-delaying, but not the phase-advancing effects of light pulses on the circadian rhythm of locomotor activity in hamsters²².

While triazolam treatment (0.5 mg) of patients suffering sleep disturbances is effective in improving sleep following a shift in the sleep-wake schedule^{23,24}, triazolam did not appear to facilitate the re-entrainment of the sleep-wake cycle to the new

schedule²⁴. However, as the results of the present study demonstrate, the phase-shifting effects of triazolam on the biological clock are dependent on the circadian time of administration. Therefore, any attempt to design an appropriate drug regime that will enhance sleep, as well as shorten the time needed for re-entrainment of the circadian system following a shift in the sleep-wake cycle, must take into account both the direction of the phase shift and the circadian time of drug administration. Drugs which can induce phase-shifts in the circadian system may be useful in the treatment of various physical and mental illnesses that have been associated with a disorder in circadian time-keeping²⁵.

This work was supported by NIH research grant HD-09885, the Whitehall Foundation and the Upjohn Company, which also provided the triazolam for this study. We thank Dr Eve Van Cauter for critical comments and Dr Kenneth Starz for assistance in the preparation of this paper.

Received 13 January; accepted 10 March 1986.

1. Akerstedt, T. *Experientia* **40**, 417–422 (1984).
2. Mellinger, G. D., Balter, M. B. & Uhlenhuth, E. H. *Archs gen. Psychiat.* **42**, 225–232 (1985).
3. Moore-Ede, M. C., Czeisler, C. A. & Richardson, G. S. *New Engl. J. Med.* **309**, 530–536 (1983).
4. Desir, D. *et al. J. clin. Endocr. Metab.* **52**, 628–641 (1981).
5. *J. Am. med. Ass.* **251**, 2410–2414 (1984).
6. Müller, W. E. in *Neuroreceptors* (ed. Hucho, F.) 3–13 (de Gruyter, Berlin, 1982).
7. Aschoff, J. & Wever, R. *Fedn Proc.* **35**, 2326–2332 (1976).
8. Czeisler, C. A., Weitzman, E. D., Moore-Ede, M. C., Zimmerman, J. C. & Kronauer, R. E. *Science* **210**, 1264–1267 (1980).
9. Weitzman, E. D. *et al. Archs gen. Psychiat.* **38**, 737–746 (1981).
10. Winfree, A. T. *Am. J. Physiol.* **245**, R497–R504 (1983).
11. Daan, S., Beersma, D. G. M. & Borbely, A. A. *Am. J. Physiol.* **246**, R161–R178 (1984).
12. Folkard, S., Hume, K. I., Minors, D. S., Waterhouse, J. M. & Watson, F. L. *Nature* **313**, 678–679 (1985).
13. Linkowski, P. *et al. J. clin. Endocr. Metab.* **61**, 429–438 (1985).
14. Wehr, T. A. & Goodwin, F. K. (eds) *Circadian Rhythms in Psychiatry* (Boxwood, Pacific Grove, California, 1983).
15. Turek, F. W. *A. Rev. Physiol.* **47**, 49–64 (1985).
16. Card, J. P. & Moore, R. Y. *Neuroscience* **13**, 415–431 (1984).
17. Van den Pol, A. N. & Tsujimoto, K. L. *Neuroscience* **15**, 1049–1086 (1985).
18. Eberts, F. S., Philopoulou, B. S., Reineke, B. S. & Vlieg, R. W. *Clin. Pharmac. Ther.* **29**, 81–93 (1981).
19. Earnest, D. J. & Turek, F. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4277–4281 (1985).
20. Ellis, G. B., McKlveen, R. E. & Turek, F. W. *Am. J. Physiol.* **242**, R44–R50 (1982).
21. Daan, S. & Pittendrigh, C. S. *J. comp. Physiol.* **106**, 253–266 (1976).
22. Ralph, M. R. & Menaker, M. *Brain Res.* **325**, 362–365 (1985).
23. Seidel, W. F., Roth, T., Roehrs, T., Zorick, F. & Dement, W. C. *Science* **224**, 1262–1264 (1984).
24. Walsh, J. K., Muehlbach, M. J. & Schweitzer, P. K. *Sleep* **7**, 223–229 (1984).
25. Van Cauter, E. & Turek, F. W. *Perspectives Biol. Med.* (in the press).

Replication of the neurochemical characteristics of Huntington's disease by quinolinic acid

M. Flint Beal, Neil W. Kowall, David W. Ellison, Michael F. Mazurek, Kenton J. Swartz & Joseph B. Martin

Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

Huntington's disease (HD) is an autosomal dominant neurological disorder characterized by progressive chorea, cognitive impairment and emotional disturbance¹. The disease usually occurs in midlife and symptoms progress inexorably to mental and physical incapacitation. It has been postulated that an excitotoxin is involved in the pathogenesis of HD^{2,3}. Schwarcz and colleagues have shown that quinolinic acid (QA) can produce axon-sparing lesions similar to those observed in HD^{4–7}. The lesions result in a depletion of neurotransmitters contained within striatal spiny neurones, for example γ -aminobutyric acid (GABA), while dopamine is unaffected. Recently, we and others have demonstrated that in HD striatum there is a paradoxical 3–5-fold increase in both somatostatin and neuropeptide Y which is attributable to selective preservation of a subclass of striatal aspiny neurones in which these peptides are co-localized^{8–12}. In the present study we demonstrate

that lesions due to quinolinic acid closely resemble those of HD as they result in marked depletions of both GABA and substance P, with selective sparing of somatostatin/neuropeptide Y neurones. Lesions produced by kainic acid (KA), ibotenic acid (IA) and *N*-methyl-D-aspartate (MeAsp) were unlike those produced by QA, as they affected all cell types without sparing somatostatin/neuropeptide Y neurones. These results suggest that QA or a similar compound could be responsible for neuronal degeneration in HD.

Although the HD gene has been localized to the short arm of chromosome 4, the pathogenesis of the disease remains unknown¹³. The excitotoxin hypothesis was advanced following the demonstration that KA lesions reproduce many of the neurochemical and neuropathological features of HD^{2,3}. These lesions, however, proved to be an imperfect model of HD as they did not spare somatostatin/neuropeptide Y neurones¹⁴⁻¹⁶. Furthermore, KA is not known to exist in mammalian tissues. It was subsequently shown by Schwarcz and colleagues that QA, an endogenous metabolite of tryptophan present in human brain, could produce excitotoxin lesions in striatum⁴⁻⁷. We therefore examined the ability of this compound to reproduce both the neurochemical and histological features of HD in rat striatum.

Male Sprague-Dawley rats (175 g) were anaesthetized with Nembutal (50 mg per kg intraperitoneal, i.p.) and positioned in a David Kopf small-animal stereotaxic apparatus. A 30-gauge blunt-tipped Hamilton syringe needle was inserted into the left striatum at the coordinates 8.4 mm anterior, 2.6 mm lateral and 4.5 mm ventral to the dura¹⁷. Drugs dissolved in 1 μ l of phosphate-buffered saline (PBS) were infused over a period of 2 min; the needle was left in place for an additional 2 min and then slowly withdrawn. Two weeks after the injection the animals were killed and their brains were removed and sectioned at 2-mm intervals as described previously¹⁰. Both striata, the frontal cortex dorsal to the lesion and hippocampus caudal to the lesion were dissected and placed in a solution of chilled 0.1 M HCl containing 1 mM EDTA and 4 mM sodium metabisulphite. The tubes containing the tissues were frozen and thawed three times, then an aliquot of supernatant was taken for analysis of catecholamines and amino acids using HPLC with electrochemical detection^{18,19}. The remaining tissue was boiled for 10 min to inactivate peptidases, and the supernatant lyophilized before reconstitution in radioimmunoassay buffer. Radioimmunoassays for somatostatin-like immunoreactivity (SLI), neuropeptide Y-like immunoreactivity (NPYLI), substance P-like immunoreactivity (SPLI) and vasopressin-like immunoreactivity (VLI) were done as described previously²⁰⁻²³. Proteins were measured by fluorimetric assay²⁴.

In an initial experiment control animals were injected with PBS alone, while experimental animals received increasing doses of QA (8 animals per dose). The right (unlesioned) striatal values did not differ from control values by one-way analysis of variance (ANOVAR); these values were therefore averaged with control values to provide a 100% control mean. The left (lesioned) striatal values were expressed as a per cent of the control mean. Control striatal concentrations of neuropeptide (per mg protein) were: SLI, 2.63 ± 0.37 pM; NPYLI, 3.47 ± 0.51 pM; SPLI, 357 ± 54 fM; GABA, 31.2 ± 5.3 nM. Increasing doses of quinolinic acid resulted in significant dose-dependent reductions in both SPLI and GABA (Fig. 1). Not until the highest dose of QA had been reached was SLI or NPYLI affected. (The highest dose was hyperosmolar and might therefore have caused nonspecific damage.) Nevertheless, even at this dose, SLI and NPYLI were preserved relative to SPLI and GABA. In contrast, increasing doses of kainate induce a graded loss of both SLI and SPLI¹¹. Dopamine concentrations were unaffected, as expected for a transmitter contained within afferent terminals. There was no significant alteration in SLI, NPYLI or SPLI in frontal cortex or hippocampus with any of the lesions.

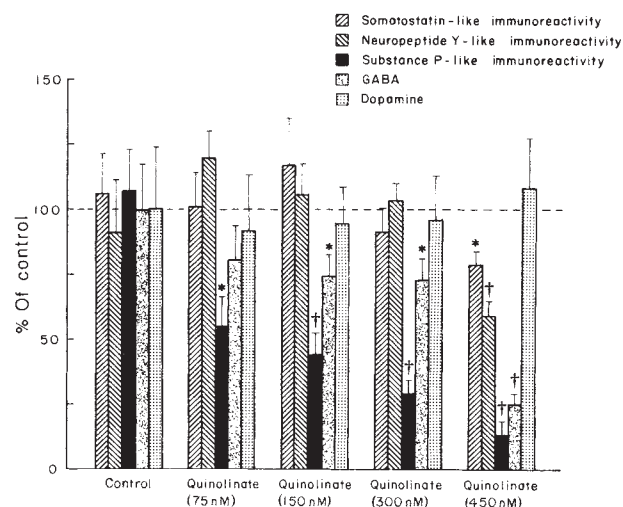


Fig. 1 Increasing doses of quinolinic acid injected into rat striatum resulted in a dose-dependent reduction in both GABA and substance P-like immunoreactivity, but somatostatin and neuropeptide Y were unaffected until the highest dose was applied. * $P < 0.05$, † $P < 0.01$ compared with controls (by ANOVAR).

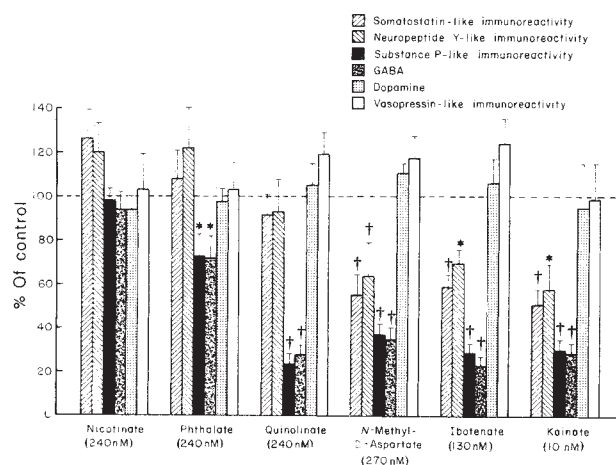


Fig. 2 Quinolinic acid (QA)-induced striatal lesions produced reductions in GABA and substance P comparable to those seen with *N*-methyl-D-aspartate, ibotenic acid and kainic acid, but only QA spared somatostatin and neuropeptide Y. Phthalic acid, which is structurally related to QA, produced similar but less potent changes than QA. * $P < 0.05$, † $P < 0.01$ compared with controls (by ANOVAR).

In a second experiment control animals received nicotinic acid (NA), while experimental animals received phthalic acid (PA), QA, MeAsp, IA or KA (Fig. 2). Nicotinic acid is structurally identical to QA except for the absence of a carboxylate group at position 2 on the pyridine ring. Eight animals in each group were killed 3 weeks after the injection and analysed as in the initial experiment. There was no significant alteration of SLI, NPYLI or SPLI in frontal cortex or hippocampus. Quinolinic acid, MeAsp, IA and KA produced comparable reductions in both SPLI and GABA but only QA lesions spared SLI and NPYLI. Dopamine and VLI (markers of striatal afferents) were unaffected by any of the lesions. Phthalic acid, which is structurally identical to QA except for a substitution of carbon for the nitrogen in the pyridine ring, was less potent than, but qualitatively similar to, QA in its toxicity. Somatostatin and neuropeptide Y were not depleted to the same extent as substance P or GABA by any of the lesions, but this probably reflects

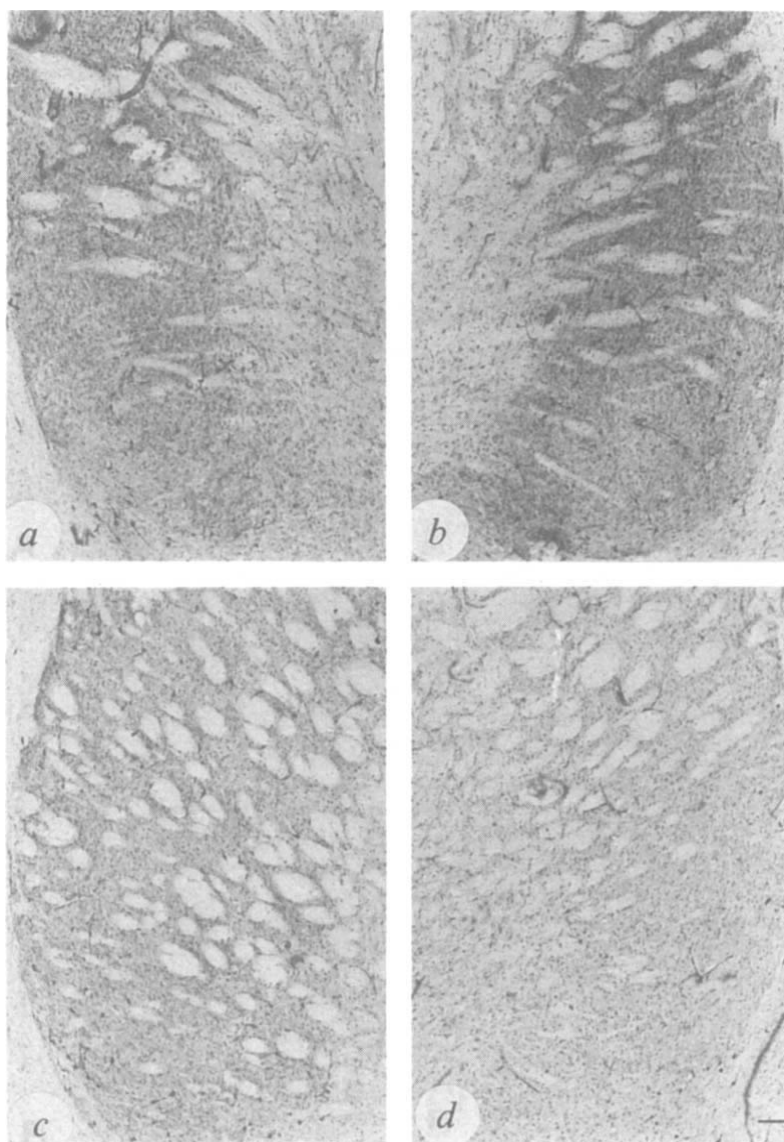


Fig. 3 *a, b*: Control (*a*; non-injected) and quinolinate-lesioned (*b*) striatum from the same section stained for NADPH-diaphorase and counterstained with cresyl violet. *c, d*: Control (*c*) and kainate-lesioned (*d*) striatum stained as for *a, b*. Both sections are equidistant from the injection sites and the lesioned sides show decreased (cresyl violet-stained) neuronal density, but this is difficult to discern because of gliosis. The diaphorase-containing neuronal population (darkly stained neurones) is resistant to the effects of quinolinate (*b*) but is depleted in concert with total neuronal loss in kainate-induced lesions (*d*).

the presence of these neuropeptides in striatal afferents as well as in intrinsic striatal neurones^{25,26}. Previous work has suggested that 25–40% of striatal somatostatin may be contained in striatal afferents^{10,25–27}.

Lesions intended for histological analysis were made identically to those in experiment 2, except that the dosage was halved by administering the drug in 0.5 μ l of saline. Five animals were studied per group. Three weeks after the lesions, the rats were pretreated with diisopropyl fluorophosphate (2 mg per kg intramuscular, i.m.) and atropine (5 mg per kg i.p.) 18 h before they were killed. Brains were immersion-fixed in 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4) at 4 °C. Sections 50 μ m thick were cut and processed for diaphorase and cholinesterase histochemistry as described previously^{28,29}. Striatal neurones that stain for diaphorase also contain both SLI and NPYLI^{30,31}.

Nicotinate injections produced no histologically identifiable striatal lesions. In contrast, quinolinate, kainate, ibotenate and phthalate all produced a zone of necrosis surrounded by an area of total neuronal loss at the injection site; neuronal counts made in adjacent sections showed neurotoxic effects, as demonstrated by decreased neuronal density. Diaphorase neurones were resistant to the neurotoxic effects of quinolinate but were more sensitive than cholinesterase neurones to the neurotoxic effects of both kainate and ibotenate (Fig. 3, Table 1). The latter

finding is consistent with a previous report³². Cholinesterase neurones were equally vulnerable to all neurotoxins examined. The histological observations indicate that preserved concentrations of both SLI and NPYLI are attributable to selective preservation of somatostatin/neuropeptide Y neurones in the face of QA neurotoxicity. Phthalic acid showed qualitatively similar effects but to a lesser degree.

As QA-induced striatal lesions spare somatostatin/neuropeptide Y neurones, they are a better model of the neurochemical and histological features of HD than lesions produced with other excitotoxins. Several other features make QA an attractive candidate for an aetiological agent of HD. It has been shown that neonatal animals are resistant to QA toxicity while mature animals are susceptible⁵; this agrees well with the typical onset of HD in middle age. In addition, it has been shown that striatum, globus pallidus and hippocampus are particularly vulnerable to quinolinic acid neurotoxicity while cerebellum, substantia nigra, amygdala, septum and hypothalamus are less susceptible⁵; many of these regions are affected in HD although the most marked pathology is seen in striatum, which suggests that other factors, such as glutamate input, may influence the topography of the pathology.

Selective sparing of somatostatin/neuropeptide Y neurones could conceivably be related to an ability of these neurones to metabolize quinolinic acid. All striatal somatostatin/neuropep-

Table 1 Quantitative analysis of diaphorase and cholinesterase neuronal populations in double-stained sections

Neurotoxin	Ratio of diaphorase cell counts (control/lesion)	Ratio of cholinesterase cell counts (control/lesion)	Fractional diaphorase/cholinesterase counts (control/lesion)
Quinolinolate	0.97 ± 0.07*	2.04 ± 0.26	0.53 ± 0.03†
Kainate	1.52 ± 0.15	1.59 ± 0.25	1.26 ± 0.10
Ibotenate	1.39 ± 0.14	1.82 ± 0.27	1.30 ± 0.17
Phthalate	1.15 ± 0.06	1.86 ± 0.27	0.71 ± 0.08

Sections at several striatal levels were counted and the cell counts on control and lesioned sides are expressed as a ratio (control striatum cell counts/lesioned striatum cell counts). A ratio of >1 indicates fewer neurones in the lesioned striatum than in the control; this is true in all cases except for diaphorase neurones in the quinolinolate-injected cases, where diaphorase cell count ratios are ~1 (that is, unchanged). This proportional sparing of diaphorase neurones is shown further by comparing the numbers of diaphorase and cholinesterase neurones (expressed as the fractional diaphorase counts, calculated by dividing the diaphorase count by the cholinesterase count) (final column). The greater number of diaphorase neurones compared with cholinesterase neurones (which are depleted) in the quinolinolate-lesioned striatum results in a ratio of <1. Ratios of >1 suggest that diaphorase neurones are more sensitive than cholinesterase neurones to kainate and ibotenate.

* The value obtained with quinolinolate is significantly less than that for kainate ($P < 0.01$) or ibotenate ($P < 0.05$) by the Kruskal-Wallis multiple-comparison test.

† The quinolinolate value is significantly less than that for kainate ($P < 0.004$) or ibotenate ($P < 0.001$) by the Kruskal-Wallis multiple-comparison test.

tide Y neurones also contain the enzyme NADPH-diaphorase^{30,31}, and it has been shown that QA can act as a precursor of NADH or NADPH in mammalian liver³³⁻³⁵. It is possible, therefore, that these neurones are able to take up and metabolize quinolinic acid; this might explain the relative sparing of the nucleus accumbens in HD^{36,37}, as concentrations of both somatostatin and neuropeptide Y as well as the density of NADPH-diaphorase neurones and terminals are 2-3-fold higher in nucleus accumbens than in the remainder of the striatum^{9,10,31}. It would also explain the paradoxical susceptibility of these neurones to MeAsp yet their resistance to QA, as both compounds are thought to act at the same receptor⁵. Some evidence, however, suggests differential effects which might implicate separate receptors^{38,39}. It has been suggested that NADPH-diaphorase neurones are resistant to a variety of neurotoxins³⁹; however, the present data suggest that this is not the case with excitotoxins, except for quinolinic acid.

Various therapeutic strategies are suggested if QA proves to be of aetiological significance in HD. Quinolinic acid neurotoxicity is dependent on cortical (presumably glutamatergic) input and can be blocked with kynurenic acid, 2-amino-5-phosphonovaleric acid and 2-amino-7-phosphonoheptanoic acid^{5,38,40,41}. It is therefore possible that early diagnosis using molecular genetics could be followed by therapeutic intervention in an attempt to retard the neurodegenerative process.

We thank Linda Lorenz for technical assistance and Ms JoAnn Forti for secretarial assistance. This work was supported by Huntington's Disease Center Without Walls (NINCDS grant NS16367) and the JulieAnne Dorn fund. M.F.B. is supported by a Young Investigator Award (IR23NS19867-1), N.W.K. and M.F.M. are Fellows of the MRC of Canada, and D.W.E. is a Fellow of the Wellcome Trust.

Received 29 January; accepted 11 March 1986.

1. Martin, J. B. *Neurology* **34**, 1059-1072 (1984).
2. Coyle, J. T. & Schwarcz, R. *Nature* **263**, 244-246 (1976).
3. McGeer, E. G. & McGeer, P. L. *Nature* **263**, 517-519 (1976).

4. Schwarcz, R., Whetsell, W. O. & Mangano, R. M. *Science* **219**, 316-319 (1983).
5. Schwarcz, R. *et al. Life Sci.* **35**, 19-32 (1984).
6. Wolfensberger, M. *et al. Neurosci. Lett.* **41**, 247-252 (1983).
7. Kohler, C., Okuno, E. & Schwarcz, R. *Neurosci. Abstr.* **11**, 820 (1985).
8. Aronin, N. *et al. Ann. Neurol.* **13**, 519-526 (1983).
9. Beal, M. F. *et al. Neurology* **34**, 663-666 (1984).
10. Nemeroff, C. B. *et al. Science* **221**, 972-975 (1983).
11. Ferrante, R. J. *et al. Science* **230**, 561-563 (1985).
12. Dawbarn, D., Dequidt, M. E. & Emson, P. C. *Brain Res.* **340**, 251-260 (1985).
13. Gusella, J. F. *et al. Nature* **306**, 234-238 (1983).
14. Beal, M. F. & Martin, J. B. *Brain Res.* **266**, 67-73 (1983).
15. Beal, M. F. *et al. Brain Res.* **361**, 135-145 (1985).
16. Allen, Y. S. *et al. Neurosci. Abstr.* **10**, 813 (1984).
17. König, J. F. R. & Klippel, R. A. *The Rat Brain* (Williams & Wilkins, Baltimore, 1963).
18. Beal, M. F. & Martin, J. B. *Neurosci. Lett.* **44**, 271-276 (1984).
19. Lasley, S. M., Greenland, R. D. & Michaelson, I. A. *Life Sci.* **35**, 1921-1930 (1984).
20. Arnold, M. A. *et al. J. Neurosci.* **2**, 674-680 (1982).
21. Beal, M. F. *et al. Neurosci. Lett.* (in press).
22. Mroz, E. A. & Leeman, S. A. in *Methods of Hormone Radioimmunoassay* (eds Jafec, B. & Behrman, J.) 121-137 (Academic, New York, 1978).
23. Kasting, N. W. *et al. Can. J. Physiol. Pharmacol.* **61**, 427-431 (1983).
24. Bohlen, P. *et al. Archs Biochem. Biophys.* **155**, 213-220 (1973).
25. Beal, M. F., Domesick, V. B. & Martin, J. B. *Brain Res.* **278**, 103-108 (1983).
26. Beal, M. F., Domesick, V. B. & Martin, J. B. *Brain Res.* **330**, 309-316 (1985).
27. Sperk, G. & Widmann, R. *J. Neurochem.* **45**, 1441-1447 (1985).
28. Kowall, N. W., Beal, M. F., Ferrante, R. J. & Martin, J. B. *Neurosci. Lett.* **59**, 61-66 (1985).
29. Vincent, S. R. *et al. Neurosci. Lett.* **35**, 111-114 (1983).
30. Vincent, S. R. *et al. J. comp. Neurol.* **217**, 252-263 (1983).
31. Kowall, N. W., Ferrante, R. J., Beal, M. F. & Martin, J. B. *Soc. Neurosci. Abstr.* **11**, 209 (1985).
32. Araki, M., McGeer, P. L. & McGeer, E. G. *Neurosci. Lett.* **53**, 197-202 (1985).
33. Henderson, L. M. & Hirsch, H. M. *J. biol. Chem.* **181**, 669-675 (1949).
34. Henderson, L. M. & Ramasarma, G. B. *J. biol. Chem.* **181**, 687-692 (1949).
35. Ghossein, R. K. *et al. J. biol. Chem.* **239**, 1208-1214 (1964).
36. Vonsattel, J. P. *et al. J. Neuropath. exp. Neurol.* **44**, 559-577 (1985).
37. McCaughey, W. T. E. *J. nerv. ment. Dis.* **133**, 91-103 (1961).
38. Foster, A. C., Vezzani, A., French, E. D. & Schwarcz, R. *Neurosci. Lett.* **48**, 273-278 (1984).
39. Perkins, M. N. & Stone, T. W. *Expl Neurol.* **88**, 570-579 (1985).
40. Thomas, E. & Pearce, A. G. E. *Acta neuropath.* **3**, 238-249 (1964).
41. Schwarcz, R. & Meldrum, B. *Lancet* **ii**, 140-143 (1985).

Channel properties of an insect neuronal acetylcholine receptor protein reconstituted in planar lipid bilayers

W. Hanke & H. Breer

Universität Osnabrück, Fachbereich Biologie/Chemie, D-4500 Osnabrück, FRG

A pentameric membrane protein composed of four types of polypeptide has been identified as the minimal structural unit responsible for the electrogenic action of acetylcholine on electrocytes and muscle cells¹⁻³. Because many populations of central and peripheral neurones also have nicotinic acetylcholine receptors (AChRs), considerable effort has recently gone into identifying the neuronal receptor^{4,5}. The central nervous tissue of insects contains very high concentrations of nicotinic AChRs^{6,7}, and we have recently purified an α -toxin binding protein, a putative AChR, from neuronal membranes of locusts^{8,9}. It is a component of high relative molecular mass, clearly composed of identical subunits, a structure predicted for an ancestral AChR protein^{10,11}. To verify that the purified polypeptides not only represent ligand binding sites but that they are indeed functional receptors, we have now reconstituted the isolated protein in a planar lipid bilayer. We show that in this system cholinergic agonists activate functional ion channels, that have properties comparable to those exhibited by the peripheral AChRs in vertebrates; thus, for the first time a functional acetylcholine receptor channel has been identified in nerve cells.

The purified receptor protein was first reconstituted into lipid vesicles produced from asolectin and cholesterol by dialysis¹². These proteoliposomes were then fused with the intact bilayer¹² to incorporate the purified receptor polypeptides. Both sides of the membrane were bathed in solution containing 100 mM NaCl, 1 mM CaCl₂ and 10 mM HEPES pH 7.2. Receptor-containing vesicles and agonists were added to the *cis* side, and membrane potentials are given with respect to virtual ground on the *trans* side.

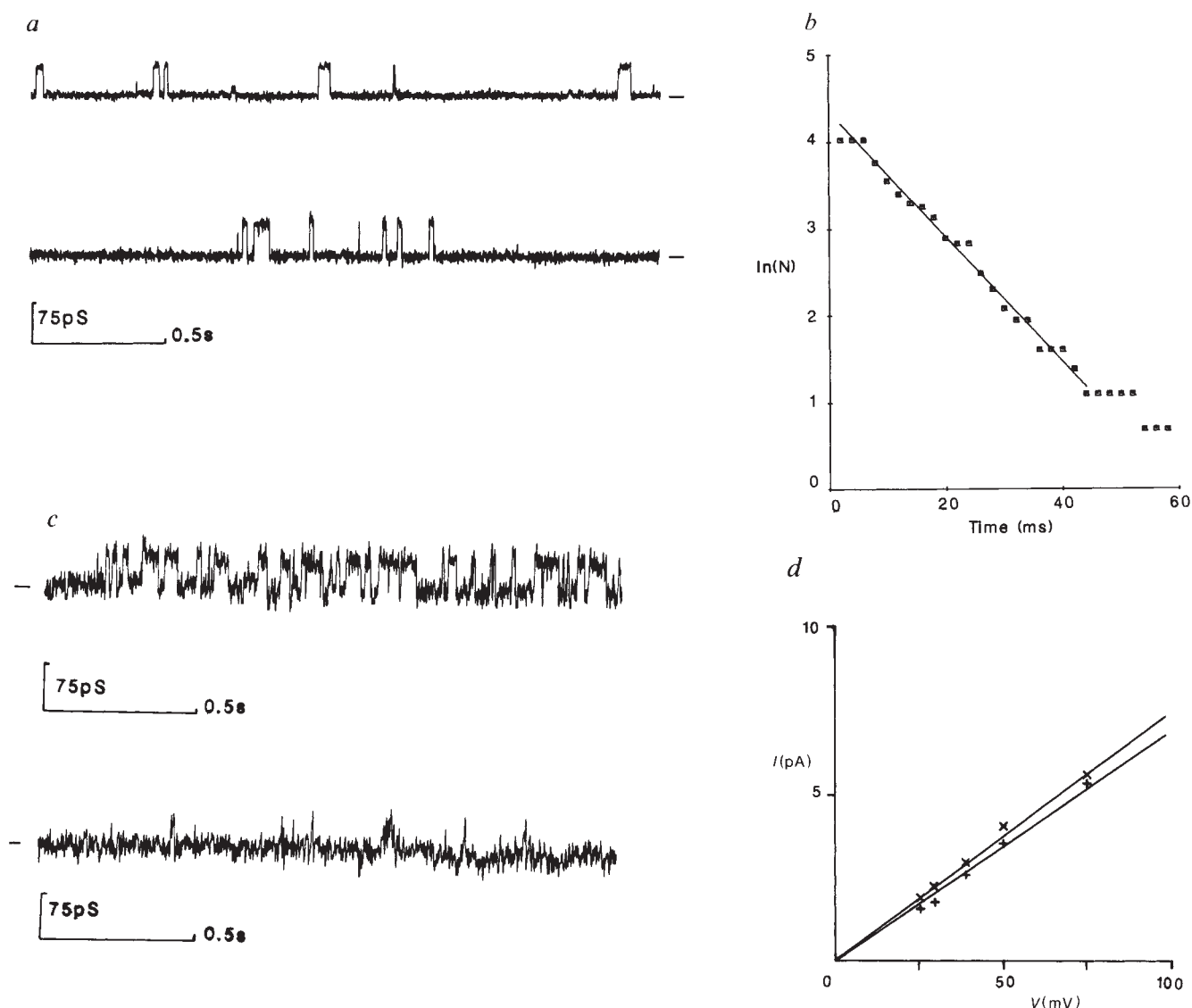


Fig. 1 Cholinergic agonist-induced fluctuations of cationic channels in planar lipid bilayers containing reconstituted α -toxin binding proteins isolated from the nervous tissue of locust. *a*, Current fluctuations of the AChR channel in the presence of $0.5 \mu\text{M}$ suberyl dicholine at an applied potential of 50 mV . *b*, Evaluation of the lifetime of current events ($n = 78$). Experimental conditions as in *a*. Mean lifetime, 14 ms . *c*, Current fluctuations of the AChR channel in the presence of two different concentrations of carbamyl choline in aqueous solution: upper trace, $50 \mu\text{M}$; lower trace, $0.5 \mu\text{M}$. Both traces were obtained from the same bilayer, starting with the lower concentration. $P(o) = 0.75$ and 0.05 in the presence of $50 \mu\text{M}$ and $0.5 \mu\text{M}$ carbamyl choline, respectively. The closed state of the channel is indicated at each trace separately (bars on left-hand side). *d*, Current-voltage ($I-V$) relation for a single channel under conditions given in *a*. $\times$, Current in the presence of 100 NaCl (75 pS); $+$, current in the presence of 100 KCl (70 pS). In these and subsequent traces, an upward deflection of the current trace corresponds to opening of a membrane channel.

Methods. Bilayers were made from 1-palmitoyl-2-oleoyl-glycero-3-phosphatidylcholine (Avanti), phosphatidylethanolamine purified from egg yolk by column chromatography in our laboratory, and dioleoyl phosphatidylglycerol (Avanti) in a $40:60:20$ molar ratio. 10 mg of lipid mixture was dissolved in 1 ml hexane and $50 \mu\text{l}$ of this solution was spread on each side of a hexadecane-pretreated hole of about $150 \mu\text{m}$ diameter. Bilayers were folded onto this hole after the solvent had been allowed to evaporate quantitatively. The lipid bilayer used was virtually solvent-free¹⁸ and generated as described previously¹²; it was shown to be electrically stable and did not respond to cholinergic agonists. Aqueous solutions on both sides of the bilayer contained 100 mM NaCl , 1 mM CaCl_2 , 10 mM HEPES at $\text{pH } 7.2$. The aqueous solutions of the vesicle preparations added were identical except that each contained additionally 200 mM sucrose. The lipid composition of the vesicles was asolectin/cholesterol in a $15:2$ molar ratio. The vesicles contained the purified native AChR protein isolated from neuronal membranes of locusts⁹. The vesicle preparation was stored on water/ice and kept there for a maximum of 3 days. $50 \mu\text{l}$ of the vesicle preparation was added to the *cis* side of the bilayer. The *trans* side was defined to be at virtual ground potential. The receptor-containing lipid bilayer was stable and no resolvable ion channel fluctuation was detected in the absence of any agonist. Agonists were added as indicated and the *cis* side was stirred until first channel fluctuations occurred. Current fluctuation traces were filtered at 300 Hz (-3 dB point, low pass) for better amplitude resolution and stored on an FM-tape recorder for later analysis. Several ($n > 5$) independent experiments were performed for each condition.

Figure 1*a* shows the current flowing through individual agonist-activated channels after addition of $0.5 \mu\text{M}$ suberyl dicholine to the *cis* side; the size of the current indicates a channel conductance of 75 pS (Fig. 1*a*, *d*). Evaluation of the current duration under these conditions (Fig. 1*b*) showed a distribution

with a single exponential indicating a mean lifetime (τ_0) of 14 ms . Figure 1*c* shows the effect of various concentrations of carbamyl choline on the same receptor channel. Again, fluctuations of a channel showing a conductance of 75 pS were observed. With $50 \mu\text{M}$ carbamyl choline, the current fluctuations

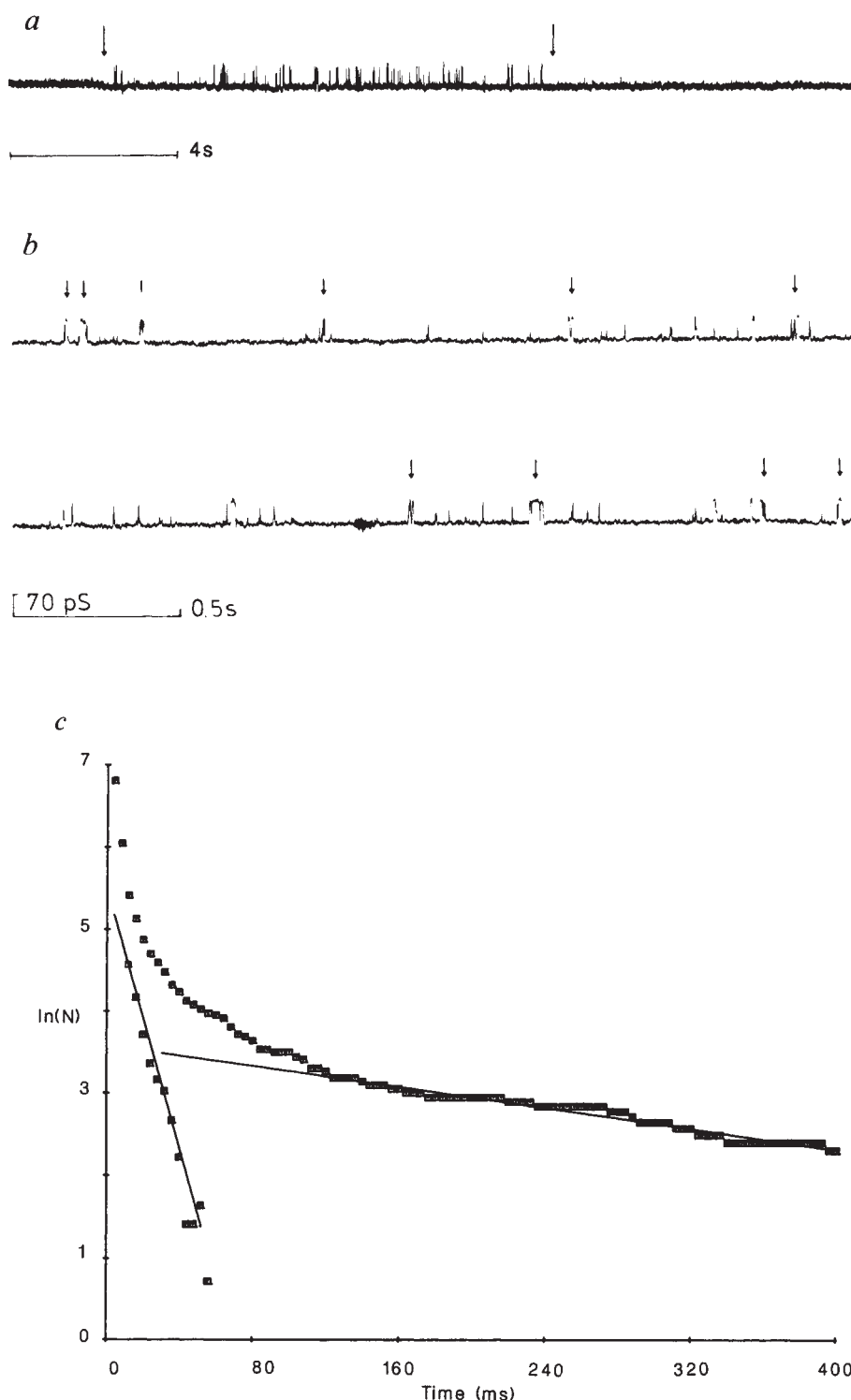


Fig. 2 Structural details of the kinetics and conductance of the neuronal AChR channel. *a*, Bursting of the channel at 25 mV. Arrows indicate the beginning and end of the burst shown. *b*, Occurrence of the multiple gating events, indicated by arrows as in *a*, on an expanded timescale. *c*, Distribution of closed time events plotted at half log scale. The graph clearly demonstrates two different components, with decay constants of $\tau_s = 315$ ms (closing time between bursts) and $\tau_f = 25$ ms (closing time within bursts). Of the 904 events evaluated, 36 were classified as belonging to the slow component. Experimental conditions were similar to those of Fig. 1, except that 100 mM NaCl was replaced by 100 mM KCl. All current fluctuations were induced by 1 μ M suberyl dicholine.

naa a mean duration of about 15 ms and a high probability of being in the open state ($P(o) = 0.75$), while with 0.5 μ M carbamyl choline the channel lifetime was only 3–5 ms and the probability of the channel being in the open state was close to zero ($P(o) \leq 0.05$). These results suggest that the α -toxin binding protein, purified from the nervous tissue of insects, behaves like a functional AChR; although quite different in its molecular organization, its response to agonist application was comparable to that of AChRs from electrocytes in lipid bilayers^{13,14}.

Comparison of agonist action at the same concentration (compare Fig. 1*a* and *c*, lower traces) shows clearly that suberyl

dicholine induced significantly greater current durations (14 ms) than did carbamyl choline (3–5 ms). A similar agonist dependency has been observed on muscle receptors¹⁵ and electrocyte receptors in lipid bilayers^{13,14}. A current-voltage curve of single channels obtained with 0.1 M NaCl on both sides of the bilayer showed that the channel behaved ohmically over the voltage range studied (± 100 mV). The conductance decreased only slightly if potassium (70 pS) was substituted for sodium (75 pS), indicating that the channel is not significantly selective for sodium over potassium (Fig. 1*d*), and conductance was unchanged when chloride was replaced by sulphate, indicating that the channel is impermeable to chloride.

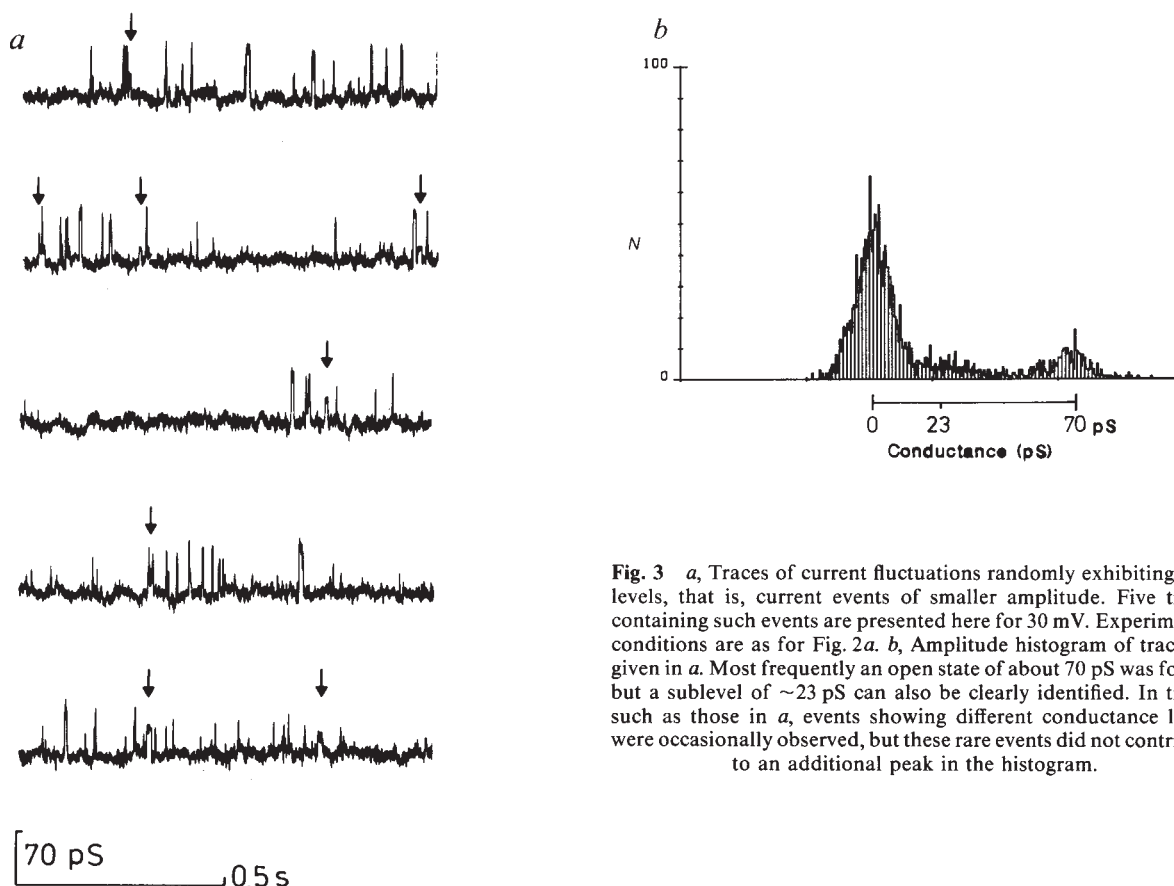


Fig. 3 *a*, Traces of current fluctuations randomly exhibiting sublevels, that is, current events of smaller amplitude. Five traces containing such events are presented here for 30 mV. Experimental conditions are as for Fig. 2*a*. *b*, Amplitude histogram of traces as given in *a*. Most frequently an open state of about 70 pS was found, but a sublevel of ~23 pS can also be clearly identified. In traces such as those in *a*, events showing different conductance levels were occasionally observed, but these rare events did not contribute to an additional peak in the histogram.

Figure 2 shows details of the neuronal receptor channel fluctuations. All current traces in this figure were obtained in the presence of 1 μ M suberyl dicholine and an aqueous solution containing symmetrical 100 mM KCl. In Fig. 2*a* the occurrence of bursts of channel activity is shown on a compressed timescale. Within these bursts channels did not open randomly, but rather, multiple gating events occurred, as shown in Fig. 2*b* on an expanded timescale. In addition to isolated single-channel openings, we found blocks of two to five or even more openings. Figure 2*c* evaluates the distribution of channel closed times under the conditions of Fig. 2*a*. The distribution is fitted by two exponentials, indicating a mean closed time between bursts (τ_s) of 300 ms and a mean closed time within bursts (τ_f) of 25 ms.

We also resolved conductance details of the ion channels created by the neuronal AChR protein (Fig. 3) and found currents indicating openings of channels with a conductance lower than the predominant channel type. These channels usually had a conductance of about 23 pS. Selected stretches of recordings show a few of these events (Fig. 3*a*), while Fig. 3*b* shows an amplitude histogram of current fluctuations to quantify this phenomenon. The agonist-induced increase in the bilayer conductance could be completely inhibited by low concentrations of (+)tubocurarine; thus, single-channel currents induced by 1 μ M suberyl choline were blocked after the addition of 5 μ M (+)tubocurarine. The antagonist did not affect the conductance of the channel, but rather reduced the frequency of occurrence of current events.

These features of the neuronal receptor—the occurrence of bursting multiple gating events and the (+)tubocurarine-induced reduction of the probability that the channel will be in the open state—have also been described for the muscle AChR of vertebrates^{16,17}. Our study shows that the purified α -bungarotoxin

binding protein isolated from insect central nervous tissue displays the properties of a chemically gated ion channel, thus representing a functional receptor. Experiments are in progress to determine the primary structure of the insect neuronal receptor protein using gas-phase microsequencing and recombinant DNA techniques; such information will allow us to estimate the structural and functional relatedness of insect and vertebrate receptors; this aspect is of particular interest in view of the fact that the complete sequence of a possible neuronal AChR polypeptide from rat has recently been reported³.

This work was supported by the Deutsche Forschungsgemeinschaft, SFB 114 and SFB 171. We thank Professor G. Boheim for helpful discussions, Miss G. Hinz for technical assistance and Dr B. Sakmann for constructive comments during the revision of the manuscript.

Received 23 December 1985; accepted 10 March 1986.

- Conti-Tronconi, B. M. & Raftery, M. A. *Rev. Biochem.* **51**, 491–530 (1981).
- Popot, J.-L. & Changeux, J. P. *Physiol. Rev.* **64**, 1162–1239 (1984).
- Maelicke, A. *Angew. Chem.* **96**, 193–219 (1984).
- Conti-Tronconi, B. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **230**, 707–717 (1985).
- Boulter, J. *et al. Nature* **319**, 368–374 (1986).
- Breer, H. *Neurochem. Int.* **3**, 43–52 (1981).
- Sattelle, D. B. *Adv. Insect Physiol.* **15**, 215–315 (1980).
- Breer, H., Kleene, R. & Benke, D. *Neurosci. Lett.* **46**, 323–328 (1984).
- Breer, H., Kleene, R. & Hinz, G. *J. Neurosci.* **5**, 3386–3392 (1985).
- Raftery, M. A., Hunkapiller, M. W., Strader, C. D. & Hood, L. E. *Science* **208**, 1454–1457 (1981).
- Noda, M. *et al. Nature* **302**, 528–532 (1983).
- Hanke, W. & Miller, C. *J. gen. Physiol.* **82**, 25–45 (1983).
- Boheim, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3586–3590 (1981).
- Labarca, P., Lindstrom, J. & Montal, M. *J. gen. Physiol.* **83**, 473–496 (1985).
- Katz, B. & Miledi, R. *J. Physiol., Lond.* **230**, 707–717 (1973).
- Neher, E. & Sakmann, B. *Nature* **260**, 799–802 (1976).
- Hamill, O. P. & Sakmann, B. *Nature* **294**, 462–464 (1981).
- Montal, M. & Mueller, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3561–3565 (1972).

Potential of synaptic transmission in the hippocampus by phorbol esters

Robert C. Malenka, Daniel V. Madison & Roger A. Nicoll

Departments of Pharmacology and Physiology, University of California, San Francisco, California 94143, USA

Protein kinase C (PKC), a calcium-dependent phospholipid-sensitive kinase which is selectively activated by phorbol esters, is thought to play an important role in several cellular processes^{1,2}. In mammalian brain PKC is present in high concentrations¹ and has been shown to phosphorylate several substrate phosphoproteins³⁻⁶, one of which may be involved in the generation of long-term potentiation (LTP)^{7,8}, a long-lasting increase in synaptic efficacy evoked by brief, high-frequency stimulation. Since the hippocampus contains one of the brain's highest levels of binding sites for phorbol esters⁹ and is the site where LTP has been most thoroughly characterized, we examined the effects of phorbol esters on hippocampal synaptic transmission and LTP. We found that phorbol esters profoundly potentiate excitatory synaptic transmission in the hippocampus in a manner that appears indistinguishable from LTP. Furthermore, after maximal synaptic enhancement by phorbol esters, LTP can no longer be elicited. Although the site of synaptic enhancement during LTP is not clearly established, phorbol esters appear to potentiate synaptic transmission by acting primarily at a presynaptic locus since changes in the postsynaptic responses to the putative transmitter, glutamate, cannot account for the increased synaptic responses induced by phorbol esters. These findings, in conjunction with previous biochemical studies, raise the possibility that, in mammalian brain, PKC plays a role in controlling the release of neurotransmitter and may be involved in the generation of LTP.

The details of the rat hippocampal slice preparation have been described elsewhere¹⁰. For all experiments, orthodromic stimulation of fibres in stratum radiatum was adjusted to evoke a 1–2-mV population spike recorded in the CA1 region of stratum pyramidale while another recording electrode placed in stratum radiatum recorded a population excitatory postsynaptic potential (e.p.s.p.) field. A variety of phorbol analogues known to activate PKC were used including 4 β -phorbol-12,13-dibutyrate (PDBu), 4 β -phorbol-12,13-diacetate (PDAc), and 12-deoxyphorbol-13-isobutyrate¹¹, all of which elicited similar responses. We also applied 4 β -phorbol and 4 α -phorbol-12,13-didecanoate, analogues which do not activate PKC¹¹. These drugs were applied by superfusion to over 50 slices. Standard techniques were used for intracellular recording.

Application of phorbol esters caused a pronounced and long-lasting increase in the amplitude of the population spike and decreased its latency (Fig. 1*a,b*). Concomitantly the initial slope and amplitude of the population e.p.s.p. increased (Fig. 1*a,b*), even at low stimulus strengths when the reflection of the population spike did not appear in the late phase of the population e.p.s.p. Washing the slice with control solution for up to 2 h did not reverse the effects of the phorbol esters. This may be due to the difficulty in completely washing out of the slice the highly lipid soluble phorbol esters. Concentrations of phorbol ester as low as 100 nM (the lowest tested) were effective although higher doses had more rapid and larger effects. Intracellular recordings corroborated these findings ($n=9$; Fig. 2*b*). The increase in the e.p.s.p. often occurred in the absence of any change in membrane potential or input resistance. Application of phorbol analogues which have been shown not to activate PKC in other systems did not enhance synaptic responses (Fig. 1*c*). This suggests that the potentiation of synaptic transmission by phorbol esters may be due to the activation of PKC, although these physiological experiments cannot entirely rule out an action of phorbol esters independent of PKC.

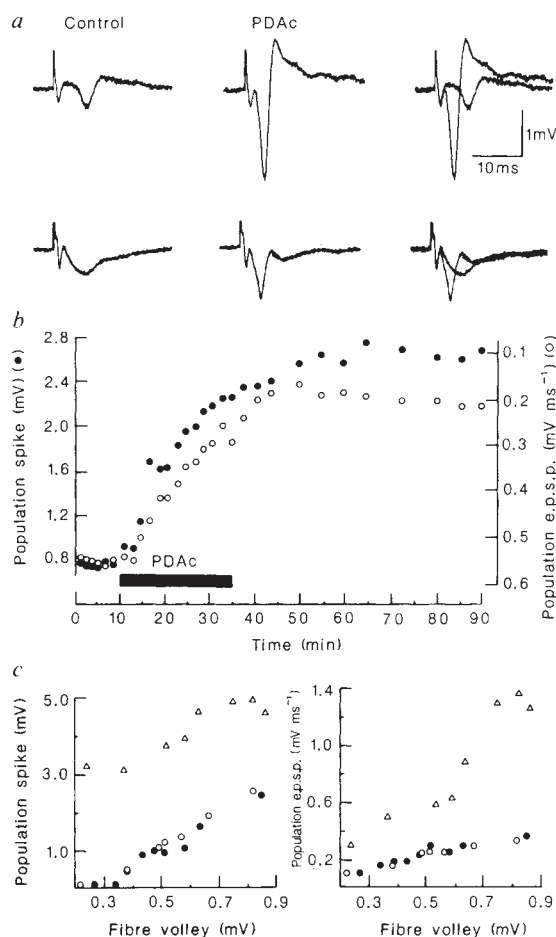


Fig. 1 Phorbol esters increase excitatory synaptic transmission in the hippocampus. *a*, The population spike (top row) and population e.p.s.p. (bottom row) are shown from an experiment in which 2 μ M phorbol-12-13-diacetate (PDAc) was applied to the hippocampus for 25 min. The responses in the PDAc column were taken 20 min after returning to drug-free solution. The responses are shown superimposed in the right-hand column. *b*, The changes in the population spike amplitude and population e.p.s.p. initial slope during this experiment; each point is the average of four responses. *c*, The population spike amplitude and population e.p.s.p. initial slope are plotted as a function of presynaptic fibre volley amplitude. Responses were obtained by varying stimulation strength in control media (●), following a 30-min exposure to a solution containing 10 μ M 4 β -phorbol (○) and also following a 20-min exposure to 5 μ M phorbol-12,13-dibutyrate (Δ), all in the same slice. Sample traces in this and all subsequent figures are averages of four responses.

Methods. The population spike corresponds to the synchronous activation of pyramidal cells. Its amplitude was determined by the method of Alger and Teyler³⁵. The initial slope of the population e.p.s.p. was measured as an indication of dendritic synaptic responses because the peak amplitude is often contaminated by the reflection of the population spike. All responses were digitized for later data analysis using a Nicolet 4094 digital oscilloscope.

One mechanism which could account for the increased size of the e.p.s.p. is that more presynaptic fibres are activated in the presence of phorbol esters than in control. To control for this variable we compared the population e.p.s.p. with the presynaptic fibre volley by constructing input-output curves (Fig. 1*c*). Application of inactive phorbol analogues had no effect on this input-output function while application of active phorbol ester to the same slice shifted this curve so that a presynaptic fibre volley of the same size elicited a much larger postsynaptic response. These results indicate that changes in presynaptic axonal excitability cannot account for the actions of phorbol esters on synaptic transmission.

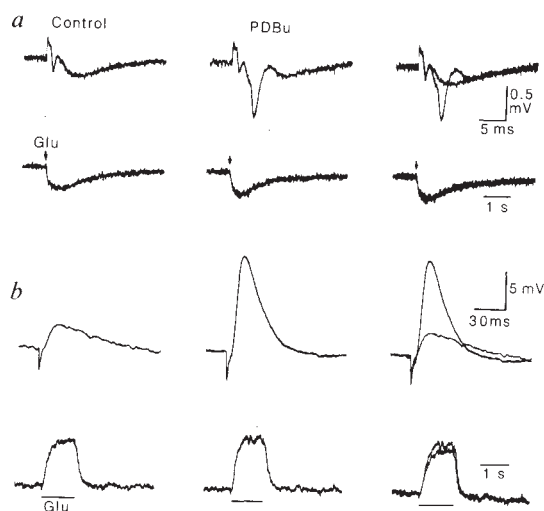


Fig. 2 Phorbol esters have minimal effects on the responses to glutamate at a time when synaptic responses are greatly increased. *a*, The population e.p.s.p. (top row) and response to pressure application of glutamate (5 mM, 40 ms, 50 p.s.i.) (bottom row) recorded with the same extracellular microelectrode are shown before and following application of 5 μ M phorbol-12,13-dibutyrate (PDBu). The responses are superimposed in the right-hand column. *b*, The intracellularly recorded e.p.s.p. (top row) and responses to iontophoretically applied glutamate (0.5 M, 20 nA) (bottom row) recorded from the same cell before and following application of 5 μ M PDBu. Responses are superimposed in the right-hand column.

Methods. For the field potential experiment, an electrode (tip diameter 2–5 μ m) attached to a Picospritzer (General Valve Corp) and filled with 5 mM glutamate in Ringer solution was positioned close to the recording microelectrode in stratum radiatum. Short pulses of glutamate (10–50 ms; 50 p.s.i.) were applied to the slice to elicit small field responses of approximately the same amplitude as the population e.p.s.p. For the intracellular experiments, standard iontophoretic techniques were used. The iontophoretic pipette was filled with 0.5 M glutamate and placed in the mid-apical dendritic zone.

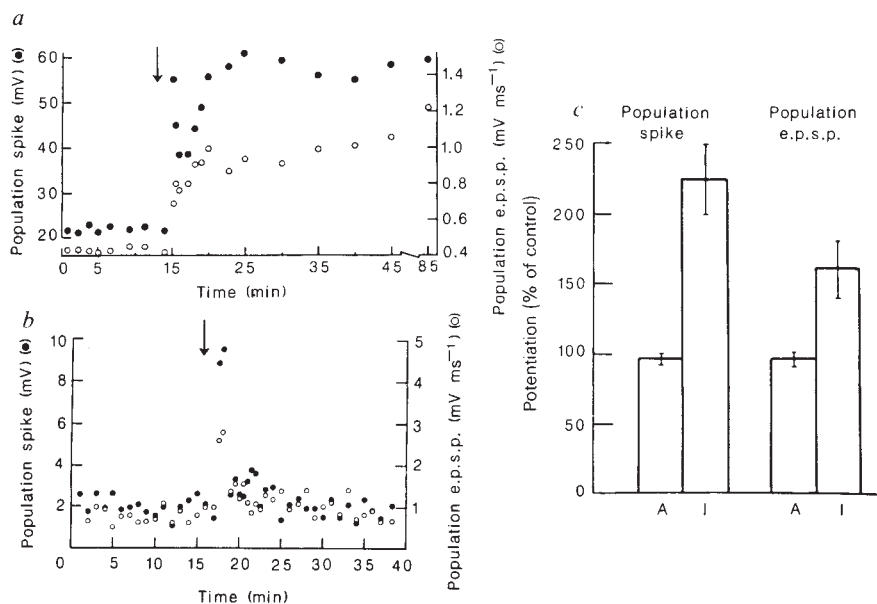
Recently, phorbol esters have been shown to have significant direct postsynaptic effects on hippocampal pyramidal cells^{12,13}. To examine the possible locus of phorbol ester's actions on synaptic transmission, we compared the effect of PDBu on the population e.p.s.p. and the field potential elicited by extracellular pressure application of glutamate¹⁴, the putative neurotransmitter at the synapse under study¹⁵. PDBu caused the usual increase in the population e.p.s.p. but failed to alter the glutamate field response (Fig. 2*a*, $n = 4$). This strongly suggests that any phorbol ester-induced change in postsynaptic parameters (for example, membrane potential) cannot account for the increase in the population e.p.s.p. While intracellularly recorded

glutamate responses were often increased, this increase could account for only a small fraction of the increase in the monosynaptic e.p.s.p. recorded from the same cell (Fig. 2*b*) (mean per cent increase in the e.p.s.p. = 135 ± 30 ($n = 4$); per cent increase for the glutamate response = 18 ± 8 ($n = 4$)). Thus it appears likely that the enhancement of synaptic transmission caused by phorbol esters is due primarily to a presynaptic action.

Comparison of synaptic enhancement during phorbol ester application and following a presynaptic tetanus which evokes LTP reveals a precise mimicry. During both conditions, the size of the population spike and e.p.s.p. increase, the population spike latency decreases, and the relationship between the fibre

Fig. 3 Increase in synaptic transmission during LTP and its blockage by phorbol esters. *a*, Changes in synaptic responses after high-frequency stimulation (arrow; two 100-Hz/1-s stimulation trains separated by 30 s) from a slice bathed for 20 min with 10 μ M of the inactive phorbol ester, 4 α -phorbol-12,13-didecanoate. Note that responses are still potentiated more than 80 min after the stimulation. *b*, The changes in synaptic responses from a slice washed for 20 min with 10 μ M PDBu are plotted following an identical stimulation as given in *a* (arrow). The responses exhibit post-tetanic potentiation but not long-term potentiation. *c*, Application of phorbol esters which activate PKC prevents LTP. Active or inactive phorbol esters were applied in a blind study. The mean ($\pm$ s.e.m.) potentiation of the population spike and e.p.s.p. from two groups of slices, one (A; $n = 10$) exposed for 20 min to 10 μ M PDBu, an activator of protein kinase C, and the other (I; $n = 10$) exposed for 20 min to 10 μ M 4 α -phorbol-12,13-didecanoate, an inert phorbol analogue. Highly significant differences between the two groups of slices (A and I) were found (population spike, $P < 0.001$; population e.p.s.p.s, $P < 0.01$; Mann-Whitney *U*-test).

Methods. For the blind study, slices were placed in the recording chamber and bathed for 20 min either with 10 μ M PDBu or with 10 μ M 4 α -phorbol-12,13-didecanoate. The experimenter did not know which randomly chosen solution had been applied until after all data analysis was completed. Only those slices meeting the minimum criterion of being able to generate a population spike of at least 5 mV were kept. Following exposure to phorbol esters, a 1–2-mV population spike was elicited every 15 s and responses recorded until a stable 20-min baseline was obtained. A tetanus was then given consisting of two, 100-Hz/1-s stimulus trains separated by 30 s using a voltage twice that used for control responses. Following the tetanus, stimulus voltage was returned to the control value and responses were recorded every 15 s for an additional 30 min. All responses taken during the 5 min prior to the tetanus were averaged and compared with the average of all responses recorded between 25 and 30 min following the tetanus. LTP was considered to have occurred if 30 min after the tetanus the population spike amplitude and/or the population e.p.s.p. initial slope were greater than 125% of control. Since the protocol for this study involved obtaining a control population spike of a fixed amplitude, lower-stimulus voltages would be used in slices exposed to PDBu. To control for this, in four additional slices we first obtained control responses before applying PDBu and for the tetanus used a stimulus voltage two times this pre-phorbol control value. Again we never observed LTP. This demonstrates that possible differences in the stimulus parameters during the tetanus cannot account for the blockade of LTP by phorbol esters.



volley and postsynaptic responses shifts in an identical manner¹⁶. If activation of PKC is involved in the generation of LTP it should be possible to occlude its generation by maximally activating this enzyme. In a pilot study we found that in seven out of seven slices treated with active phorbol esters we were unable to generate LTP while in seven out of eight control slices LTP was elicited. Based on these findings, we undertook a more rigorous blind study (see Fig. 3 legend) to test this hypothesis. Figure 3 compares the results of high-frequency stimulation in two slices from the blind study, one bathed in an inactive phorbol analogue (Fig. 3a) and the other bathed in PDBu (Fig. 3b). When stimulated with a tetanus, slices bathed in the inactive phorbol exhibited a brief post-tetanic potentiation (PTP) followed by LTP (Fig. 3a), while slices bathed in PDBu exhibited only PTP (Fig. 3b). Figure 3c summarizes the findings of the blind study. Of the 10 slices exposed to the inactive phorbol analogue, 8 exhibited LTP whereas none of the 10 slices exposed to PDBu exhibited LTP. However, since PDBu did not prevent PTP (Fig. 3b) the actions of phorbol esters appear to be specific to the long-lasting form of synaptic potentiation. Consistent with our results is the recently reported preliminary finding that phorbol esters prolong the decay of LTP *in vivo*¹⁷.

The mechanisms underlying synaptic plasticity have generated a great deal of discussion. The most common site responsible for alterations in synaptic strength has been reported in a variety of systems to be the presynaptic terminal¹⁸⁻²¹. Based on our experiments examining glutamate responses and input-output curves, we propose that the enhancement of synaptic transmission induced by phorbol esters is also caused by an action localized to the presynaptic terminal. Consistent with this proposal are several biochemical studies performed in brain tissue which have demonstrated that PKC is present in presynaptic nerve terminals²², is activated during depolarization of the presynaptic nerve terminal³, and phosphorylates proteins found in presynaptic terminal membrane³⁻⁶. Furthermore, phorbol esters have been shown to enhance calcium-dependent neurotransmitter release from both rat brain synaptosomes²³ and cultured brain neurones²⁴.

The precise mechanism underlying the enhancement of neurotransmitter release induced by phorbol esters is unknown. Phorbol esters have been shown to block a calcium-activated potassium current in hippocampal pyramidal cells^{12,13} and similar channels have been shown to exist in rat presynaptic nerve terminals²⁵. Blockage of these channels might prolong the action potential in the terminal, thus increasing transmitter release. A similar mechanism involving a different type of potassium channel has been proposed to account for the enhanced transmitter release which underlies behavioural sensitization in the marine mollusc, *Aplysia*²⁶. Phorbol esters have also been shown to directly enhance calcium currents in an invertebrate neurone²⁷. Alternatively, phorbol esters may sensitize the secretory process to calcium as has been shown in other cell types². A presynaptic locus may also account for the generation of LTP since increases in glutamate release have been demonstrated during LTP in the hippocampus^{28,29}. However there are also several experiments consistent with significant postsynaptic modifications^{30,31}.

Several proteins have been shown to be phosphorylated following the generation of LTP in the hippocampus^{7,8,32,33}. One of these, Fl (possibly the same as the synaptically associated B50), also appears to be a substrate for PKC^{7,8} although its anatomical localization and function are as yet unknown. If PKC activation is necessary for the induction of LTP, one possible sequence of events for the induction of LTP could be that the high-frequency stimulation raises intracellular calcium concentration to a level that results in a significant activation of PKC. In fact, a recent report has demonstrated that in erythrocyte vesicles small increases in calcium concentration can cause a translocation of PKC from a soluble to a membrane-bound compartment³⁴. An alternative hypothesis would be that

high-frequency stimulation causes the release of some extracellular messenger capable of stimulating inositol phospholipid turnover. This would result in the generation of diacylglycerol and the subsequent activation of PKC.

This work was supported by NIH grants MH-38256, RCDA MH-00437, the Klingenstein Fund (to R.A.N.), NS-07495 (to R.C.M.) and the Scottish Rite Schizophrenia Research Program NMJ, USA.

Note added in proof: Recently a report has appeared³⁶ demonstrating translocation of PKC from soluble to membrane fractions following LTP in the hippocampus, thus supporting a role for PKC in LTP.

Received 7 November 1985; accepted 6 March 1986.

- Nishizuka, Y. *Nature* **308**, 693-698 (1984).
- Baker, P. F. *Nature* **310**, 629-630 (1984).
- Wu, W. C.-S., Walaas, S. I., Nairn, A. C. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5249-5253 (1982).
- Sorensen, R. G., Kleins, L. P. & Mahler, H. R. *Brain Res. Bull.* **7**, 57-61 (1981).
- Kristjansson, G. I., Zwiers, H., Oestreicher, A. B. & Gispen, W. H. *J. Neurochem.* **39**, 371-378 (1982).
- Aloyo, V. J., Zwiers, H. & Gispen, W. H. *J. Neurochem.* **41**, 649-653 (1983).
- Nelson, R. B. & Routtenberg, A. *Expl. Neurol.* **89**, 213-224 (1985).
- Akers, R. F. & Routtenberg, A. *Brain Res.* **334**, 147-151 (1985).
- Murphy, K. M., Gould, R. J., Oster-Granite, M. L., Gearhart, J. D. & Snyder, S. H. *Science* **222**, 1036-1038 (1983).
- Alger, B. E. & Nicoll, R. A. *J. Physiol., Lond.* **328**, 105-123 (1982).
- Castagna, M. *et al. J. biol. Chem.* **257**, 7847-7851 (1982).
- Baraban, J. M., Snyder, S. H. & Alger, B. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2538-2542 (1985).
- Malenka, R. C., Madison, D. V., Andrade, R. & Nicoll, R. A. *J. Neurosci.* **6**, 475-480 (1986).
- Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132-137 (1984).
- Cotman, C. W. & Nadler, J. V. in *Glutamate: Transmitter in the Central Nervous System* (eds Roberts, P. J., Storm-Mathisen, J. & Johnston, G. A. R.) 117-154 (Wiley, Chichester, 1981).
- Andersen, P., Sundberg, S. H., Sveen, O., Swann, J. W. & Wigström, H. *J. Physiol., Lond.* **302**, 463-482 (1980).
- Lovinger, D., Colley, P., Linden, D. J., Murakami, K. & Routtenberg, A. *Soc. Neurosci. Abstr.* **11**, 927 (1985).
- Kandel, E. R. *Nature* **293**, 697-700 (1981).
- Briggs, C. A., McAfee, D. A. & McCaman, R. E. *J. Physiol., Lond.* **363**, 181-190 (1985).
- Koyano, K., Kuba, K. & Minota, S. *J. Physiol., Lond.* **359**, 219-233 (1985).
- Baxter, D. A., Bittner, G. D. & Brown, T. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5978-5982 (1985).
- Girard, P. R., Mazzei, G. J., Wood, J. G. & Kuo, J. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3030-3034 (1985).
- Nichols, R. A., Haycock, J. W., Wang, J. K. T. & Greengard, P. *Soc. Neurosci. Abstr.* **11**, 846 (1985).
- Zurgil, N. & Zisapel, N. *FEBS Lett.* **185**, 257-261 (1985).
- Bartschat, D. K. & Blaustein, M. P. *J. Physiol., Lond.* **361**, 441-457 (1985).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).
- DeRiemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. K. *Nature* **313**, 313-316 (1985).
- Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496-498 (1982).
- Bliss, T. V. P., Douglas, R. M., Errington, M. L. & Lynch, M. J. *J. Physiol., Lond.* **361**, 50P (1985).
- Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. & Schottler, F. *Nature* **305**, 719-721 (1983).
- Wigström, H. & Gustafsson, B. *Nature* **301**, 603-604 (1983).
- Browning, M., Dunwiddie, T., Bennett, W., Gispen, W. & Lynch, G. *Science* **203**, 60-62 (1979).
- Bär, P. R., Schotman, P., Gispen, W. H., Tielen, A. M. & Lopes da Silva, F. H. *Brain Res.* **198**, 478-484 (1980).
- Wolf, M., LeVine, H. III, May, W. S. Jr, Cuatrecasas, P. & Sahyoun, N. *Nature* **317**, 546-549 (1985).
- Alger, B. E. & Teyler, T. J. *Brain Res.* **110**, 463-480 (1976).
- Akers, R. F., Lovinger, D. M., Colley, P. A., Linden, D. J. & Routtenberg, A. *Science* **231**, 587-589 (1986).

Platelet activation—a role for a 40K anti-phospholipase A₂ protein indistinguishable from lipocortin

Lhousseine Touqui, Bernard Rothhut, Angus M. Shaw, Armel Fradin, B. Boris Vargaftig & Françoise Russo-Marie

Unité associée Institut Pasteur/Inserm No. 285, Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, France

Stimulus-response (S-R) coupling in platelets requires an intermediary other than an elevation in cytosolic free calcium ([Ca²⁺]_i). While an increase in [Ca²⁺]_i is essential in S-R coupling, effecting phosphorylation of myosin of relative molecular mass (M_r) 20,000 (20 K)³, platelet activation is also associated

with phosphorylation of a 40K protein⁴, which can occur in the absence of changes in $[Ca^{2+}]_i$ (ref. 5). The 40K protein is the substrate for protein kinase C (PKC)^{6,7}. Mounting evidence suggests that activation of PKC by diacylglycerol is the other signal involved in S-R coupling⁸⁻¹⁰. Although phosphorylation of the 40K protein is associated with certain platelet functional responses^{4,6,8,11,12}, no precise role has been accredited to it. Recently, we and others have described several proteins (collectively known as lipocortin) which inhibit phospholipase A₂ (PLA₂)¹³⁻¹⁹. One of the most conspicuous proteins of this group is a 40K peptide whose inhibitory activity can be suppressed by prior phosphorylation^{18,20}. We hypothesized that the 40K protein described in platelets may possess anti-PLA₂ activity and that phosphorylation by PKC, suppressing its inhibitory activity, may represent the mechanism underlying mobilization of arachidonic acid, the precursor of prostaglandins. The results of the present study strongly support this hypothesis.

Agonist-induced platelet activation is closely associated with phosphorylation of a 40K protein^{4,6,8,10-12} which is the substrate for PKC. Although activators of PKC such as phorbol myristate acetate (PMA) or 1-oleyl-2-acetylgllycerol (OAG) can induce some platelet responses, such as aggregation and degranulation⁹, no specific function has been reported for the 40K protein. Recent studies have shown that activators of PKC, such as PMA, enhance the activity of PLA₂ in calcium ionophore-stimulated platelets²¹, suggesting a possible role for PKC/40K protein in the regulation of PLA₂ activity. Further support for such a role

derives from studies¹⁸ which demonstrated, in peritoneal neutrophils, the existence of an anti-phospholipase A₂ protein with a M_r of 40K whose anti-PLA₂ activity can be reduced by prior phosphorylation. The phosphorylation of this protein, originally named lipomodulin and now termed lipocortin¹⁶, in neutrophils parallels the release of arachidonic acid during stimulation by a chemo-attractant. Subsequent investigations²⁰ demonstrated that in murine thymocytes this protein is phosphorylated by a PKC-dependent mechanism.

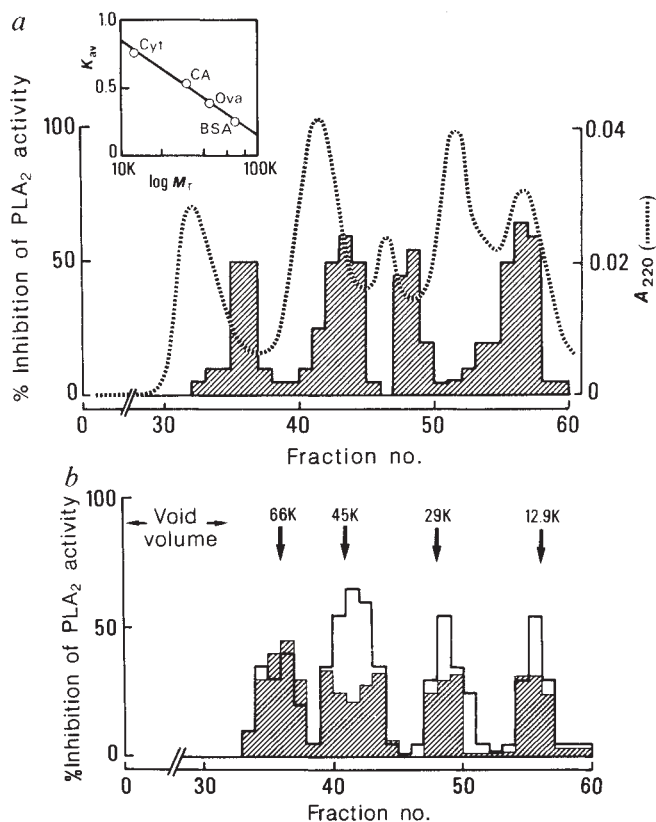
On the basis of these observations, we hypothesize that the platelet 40K protein may have anti-PLA₂ activity and that this activity may be regulated by phosphorylation, in a manner similar to that described for lipocortin in neutrophils¹⁸. Thus, by altering the anti-PLA₂ activity of the 40K protein, activation of PKC may have an important role in the regulation of PLA₂ activity in platelets. To substantiate this hypothesis, we extracted platelet proteins before and after activation of PKC and assayed the extracted proteins for PLA₂ inhibitory activity.

Plasma-free platelet suspensions were prepared from whole blood obtained from the central ear artery of adult New Zealand White rabbits, as described elsewhere²². Platelets were lysed by glycerol loading²³ and homogenized in the presence of a membrane solubilizer, Tween-20 (0.5% v/v). Proteins were separated by HPLC using a TSK G2000 SW Ultropac column. The mobile phase, adjusted to a flow rate of 0.2 ml min⁻¹, consisted of Tris-NaCl (0.1–0.15 M, pH 7) containing 2 mM EDTA. The activity of each HPLC fraction (0.4 ml) was assessed by monitor-

Fig. 1 Anti-phospholipase A₂ activity of HPLC fractions. **a**, Estimation of the M_r s of inhibitory fractions from untreated platelets.

b, Anti-PLA₂ activity of fractions from thrombin-stimulated platelets. Washed platelets (2×10^9 cells ml⁻¹) were suspended in Tyrode's-gelatin (TG) buffer containing (1 mM) MgCl₂, 2 mM CaCl₂ and 0.25% (w/v) gelatin (pH 7.4).

a, Aliquots (3 ml) of the platelet suspension were lysed by glycerol loading, then 1 ml of Ca²⁺-free TG buffer was added to each aliquot to induce osmotic shock. Lysates were then homogenized in the presence of a membrane solubilizer, Tween 20 (0.5% v/v). Ca²⁺-free TG buffer was added to the homogenates to attain a final protein concentration of 4 mg ml⁻¹. These mixtures were stored in aliquots at -60 °C until required. Aliquots of homogenates containing 0.2 mg protein were subjected to HPLC and the proteins separated on a TSK G 2000 SW Ultropac column using Tris-NaCl buffer (0.1–0.15 M, pH 7) as the mobile phase. The M_r s of the resolved proteins were determined by referring to a calibration curve (inset) derived using protein standards [bovine serum albumin (BSA), ovalbumin (Ova), carbonic anhydrase (CA) and cytochrome C (Cyt)]. The anti-PLA₂ activity of each fraction was assessed by monitoring its ability to inhibit the release of porcine pancreatic PLA₂-induced ³H-oleic acid from a suspension of labelled *E. coli* membranes³¹. Briefly, 0.1-ml aliquots from each HPLC fraction were each added to 0.25 ml of Tris buffer (100 mM, pH 8) containing 10 mM CaCl₂ and preincubated with porcine pancreatic PLA₂ (50 ng) for 10 min at 4 °C. In controls, protein fractions were substituted with HPLC buffer. Reactions were initiated by adding *E. coli* substrate (25 µl, 250,000 c.p.m.) and terminated 5 min later by the addition of 50 µl 2 M HCl. The released oleic acid was trapped in the supernatant by the addition of excess BSA (100 mg ml⁻¹), and after centrifugation (10,000g, 5 min) to remove particulate material, an aliquot of supernatant was counted for radioactivity by liquid scintillation spectrophotometry. In the absence of any inhibitor, 50 ng PLA₂, over a period of 5 min, induced the release of 8–12% of incorporated oleic acid, a value which represented between 40 and 60% of the maximum response. The inhibition of the anti-PLA₂ activity of the various fractions from one typical experiment is expressed as a percentage of the release evoked in the absence of an inhibitor. The dotted line, representing the profile of the resolved proteins, was determined by measuring the optical absorbance of the eluate at 220 nm. **b**, Aliquots (3 ml) of the platelet suspension were challenged with thrombin (1 U ml⁻¹) (or saline in the case of controls) for 1 min at 37 °C. Reactions were stopped by adding ice-cold EDTA (5 mM final concentration). Thereafter, cells were lysed, homogenized and anti-PLA₂ activity determined (following lysis control cells received 1 U ml⁻¹ thrombin). The cumulative results from five experiments are shown; error bars have been omitted for clarity. Compared with controls (open blocks), thrombin treatment (hatched blocks) decreased the inhibitory activity of the fractions containing the 40K, 30K and 15K proteins by 54.7 ± 5.2% ($P < 0.001$), 31.0 ± 5.5% ($P < 0.01$) and 10.0 ± 2.2% (not significant), respectively (mean ± s.e.m.). When thrombin was replaced by PMA (10 nM) a marked decrease in the anti-PLA₂ activity of the 40K fraction was observed (see text).



ing its ability to inhibit porcine pancreatic PLA₂-induced ³H-oleic acid release from a suspension of labelled *Escherichia coli* membranes (see Fig. 1 legend).

Four fractions, with apparent *M_s* of 15K, 30K, 40K and 60K, were found to possess anti-phospholipase activity (Fig. 1a), suggesting that rabbit platelets contain proteins which are similar to those now referred to as lipocortin¹⁶. The principle *M_s* reported for the lipocortin proteins are 15K, 30K and 40K. The inhibition associated with the 60K fraction has been reported previously¹⁷ and appears to be due to a nonspecific effect of albumin. The 30K and 15K peptides are considered to be active proteolytic fragments of the 40K protein as in the presence of protease inhibitors (which were omitted in the present study) only the 40K fraction is recovered from the cells²⁴.

To determine whether the 40K protein extracted from rabbit platelets was similar to lipocortin, we examined the effect of a monoclonal antibody (BF26)²⁵ raised against renocortin—a lipocortin from rat renomedullary interstitial cells^{15,25}—on the ability of the protein fractions to inhibit PLA₂ activity in an *in vitro* assay (see Fig. 1 legend). Fractions containing the 40K protein were incubated with various dilutions of antibody for 1 h at 37 °C and for 24 h at 4 °C. The anti-PLA₂ activity following incubation with a 1:50 dilution of BF26 was reduced from 28.3 ± 3.5% to 6.0 ± 0.7% (mean ± s.e.m. of six observations, *P* < 0.001). Controls, in which antibody alone was incubated with HPLC buffer, had no effect on PLA₂ activity. This result demonstrates that the 40K protein is immunologically identical to lipocortin.

We further questioned the role of lipocortin in platelets by observing the effect of thrombin, a potent activator of PKC⁴, on the anti-phospholipase activity of the extracted proteins. Following treatment of intact platelets with this agonist, the inhibitory activity of the 40K, 30K and 15K fractions was reduced (Fig. 1b), the predominant effect being observed with the 40K fraction; no effect on the 60K fraction was observed. In addition, when these fractions were pretreated with alkaline phosphatase, their anti-phospholipase activities were restored (Table 1), suggesting that the effect of thrombin on lipocortin

Table 1 Effect of alkaline phosphatase on HPLC fractions from thrombin-treated platelets

Fraction	% Inhibition of PLA ₂ activity	
	Control	Alkaline phosphatase-treated
40K	38.9 ± 7	64.3 ± 4.0*
30K	22.8 ± 6.5	42.8 ± 2.9*
15K	21.7 ± 5.1	51.3 ± 5.1*

The homogenate from thrombin-treated platelets was subjected to HPLC and 0.3-ml aliquots from fractions corresponding to *M_s* of 40K, 30K and 15K were each incubated with 1.2 U of insoluble alkaline phosphatase (AP) (*pH* 10) at 20 °C for 90 min. The enzyme was then recovered by centrifugation (5,000g, 5 min) and the anti-PLA₂ activity of these and time-matched controls, which received AP buffer alone, was assayed as already described. Values are the mean ± s.e.m. of 7–8 observations.

* Significant difference compared with the control (*P* < 0.001).

(that is, a reduction of its anti-PLA₂ activity) was linked to a phosphorylating event. This notion was subsequently confirmed using ³²P-labelled platelets treated with thrombin or PMA and subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE). There was increased incorporation of ³²P in two major bands, one corresponding to myosin light chain (20K) and another corresponding to 40K (Fig. 2b,c). To establish whether the 40K protein was in fact lipocortin, aliquots of homogenates from thrombin- or PMA-stimulated, ³²P-labelled platelets were incubated with the monoclonal antibody BF26. Immunoprecipitated proteins were pelleted and subjected to SDS-PAGE; the

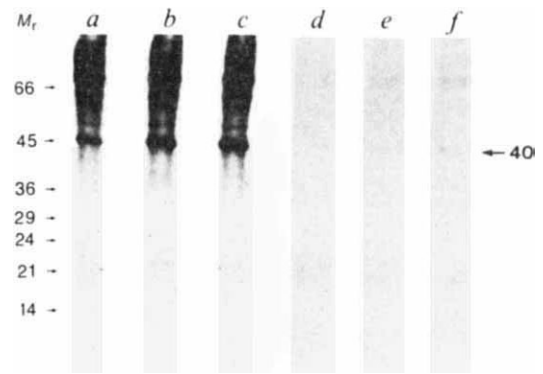


Fig. 2 Immunoprecipitation and SDS-PAGE of ³²P-labelled proteins from platelets. The figure shows the autoradiogram of proteins from control (a), thrombin-treated (b) and PMA-treated (c) platelets before immunoprecipitation. d, Immunoprecipitation of proteins from thrombin-treated platelets by nonspecific antibody. e, f, Immunoprecipitation with specific antibody (BF26) of proteins from thrombin(e)- or PMA(f)-treated platelets. The arrow indicates the 40K protein immunoprecipitated by the specific antibody. **Methods.** Washed platelets were incubated with ³²P (0.5 mCi ml⁻¹) for 1 h at 22 °C and aliquots (100 µl) were incubated with thrombin (1 U ml⁻¹), PMA (10 nM) or vehicle alone. After 1 min, reactions were stopped by adding 50 µl of RIPA buffer (*pH* 7.2) containing 140 mM NaCl, 10 mM Tris-HCl, 10 mM EDTA, 1% Nonidet P-40, 0.1% SDS, 1 mM phenylmethylsulphonyl fluoride, 200 IU ml⁻¹ aprotinin, 30 mM nitrilotriacetic acid and 30 mM sodium fluoride. Aliquots (15 µl, each containing 100,000 c.p.m.) of homogenates obtained from thrombin- or PMA-treated platelets were incubated at 4 °C with protein A-Sepharose (60 mg ml⁻¹) for 30 min and nonspecific protein binding was subsequently eliminated by centrifugation (5 min at 5,000g). Then, 90 µl of supernatant were incubated with anti-renocortin antibody BF26 (26.7 mg protein ml⁻¹) at 37 °C. In the case of the control, aliquots (90 µl) from thrombin-treated platelets received an equivalent amount of non-specific antibody of the same class (IgG1) as BF26. After 1 h, rabbit anti-mouse antiserum (1:10 dilution) was added to increase specific binding; 30 min later, protein A-Sepharose (60 mg ml⁻¹) was added and the mixture incubated for 18 h at 4 °C. The antibody-bound proteins were pelleted by centrifugation then subjected to SDS-PAGE as described in ref. 32 and radioactivity visualized by autoradiography. Total ³²P incorporated into proteins, determined by trichloroacetic acid precipitation, was increased 2–3-fold in thrombin-treated platelets. The radioactivity immunoprecipitated by BF26 (316 ± 38 c.p.m.) in proteins from control platelets increased by 3.7-fold (to 1,168 ± 160 c.p.m.) in proteins from thrombin-treated platelets. In contrast, no significant radioactivity was immunoprecipitated by the nonspecific antibody. Similar results were obtained with proteins from PMA-treated platelets (data not shown).

resultant autoradiograms showed a distinct band corresponding to a *M_r* of 40K (Fig. 2e,f), demonstrating that at least part of this protein is immunologically identical to lipocortin. The 30K and 15K fractions (proteolytic fragments of 40K) did not appear in the autoradiograms, as protease inhibitors were present in these experiments.

The present study shows that rabbit platelets contain a 40K anti-phospholipase protein which is recognized by a specific anti-lipocortin antibody (from renomedullary cells), indicating that this protein is immunologically indistinguishable from lipocortin. As shown in Fig. 2, the radioactivity associated with the immunoprecipitated band does not represent all of the phosphorylated 40K protein observed after stimulation with thrombin or PMA; one possible explanation for this is that the 40K protein is not unique but consists of various different proteins, as reported by Imaoka *et al.*⁷, and that lipocortin is one of these proteins.

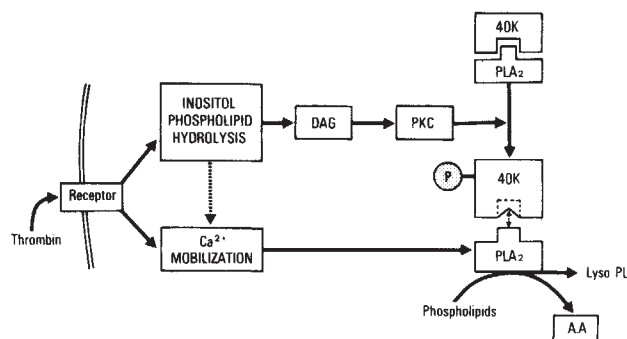


Fig. 3 Synergistic action of protein kinase C (PKC) activation and Ca^{2+} mobilization for eliciting full PLA_2 activation in thrombin-stimulated platelets.

Previous studies²⁰ have shown that the anti- PLA_2 activity of lipocortin from murine thymocytes can be regulated by PKC-dependent phosphorylation. Here we have shown that, in response to thrombin, the anti- PLA_2 activity of the 40K protein is decreased and that this reduction appears to be associated with phosphorylation of the 40K protein. A role for PKC in suppressing the anti- PLA_2 activity of the 40K protein is therefore suggested. Indeed, when thrombin was replaced by 10 nM PMA, a specific activator of PKC, the anti- PLA_2 activity of the 40K protein was reduced from $71.6 \pm 9.9\%$ in dimethyl sulphoxide (DMSO)-treated control platelets to $16.6 \pm 10.6\%$ after PMA treatment (mean $\pm$ s.e.m. of eight observations $P < 0.01$), further suggesting that the anti- PLA_2 activity of the 40K protein in platelets is regulated by PKC.

It has been proposed²⁶ that two intermediaries—diacylglycerol (DAG) and an elevation in $[\text{Ca}^{2+}]_i$ —are involved in S-R coupling in various cell types. Here we propose that phospholipase A_2 is also regulated by these two intermediaries to achieve optimal activity. The regulation of PLA_2 by Ca^{2+} and DAG is summarized in Fig. 3. In this scheme, receptor activation leads to an increase in $[\text{Ca}^{2+}]_i$ and the concomitant formation of DAG, from the hydrolysis of inositol phospholipids²⁶. Activation of PKC, by DAG, phosphorylates the 40K lipocortin, suppressing its anti- PLA_2 activity. In the presence of an agonist-induced elevation in cytosolic $[\text{Ca}^{2+}]_i$ (PLA_2 enzymes have an absolute dependency on ionic Ca^{2+} ; ref. 27) optimal activity is evoked.

In preliminary experiments, we have observed that the PKC activator PMA, at concentrations that induce phosphorylation of the 40K protein, does not cause the release of ^3H -arachidonic acid (^3H -AA) from labelled platelets. However, the release of AA induced by the Ca^{2+} ionophore A23187 (250 nM) was enhanced from $15 \pm 4\%$ to $38 \pm 5\%$ (mean $\pm$ s.e.m. of four experiments, $P < 0.05$) by pretreating the platelets with PMA (10 nM), which further lends credence to our proposed scheme. A similar PMA-mediated potentiation of ionophore-induced AA release and of thromboxane B_2 formation in rabbit neutrophils and human platelets has been reported elsewhere^{21,28}. More recently, both PMA and DAG have been shown to potentiate the release of AA from ionophore-stimulated human platelets²⁹.

The nature of the interaction between lipocortin and PLA_2 has not yet been elucidated. In Fig. 3 we have shown lipocortin, in its unphosphorylated state, bound to and inactivating PLA_2 , rather like the non-enzymatic subunit of crotoxin in snake venom³⁰. Phosphorylation of lipocortin dissociates the complex, allowing, in the presence of an elevation in cytosolic calcium, optimal activity of PLA_2 .

- Lyons, R. M., Stanford, N. & Majerus, P. W. *J. clin. Invest.* **56**, 924–936 (1975).
- Yamanishi, J. *et al. Biochem. biophys. Res. Commun.* **112**, 778–786 (1983).
- Sano, K., Takai, Y., Yamanishi, J. & Nishizuka, Y. *J. biol. Chem.* **258**, 2012–2013 (1983).
- Imaoka, T., Lynham, J. A. & Haslam, R. J. *J. biol. Chem.* **258**, 11404–11414 (1983).
- Kajikawa, N. *et al. Biochem. biophys. Res. Commun.* **116**, 743–750 (1983).
- Rink, T. J., Sanchez, A. & Hallam, T. J. *Nature* **305**, 317–319 (1983).
- Haslam, R. J. & Davidson, M. L. *Biochem. J.* **222**, 231–261 (1984).
- Haslam, R. J. & Lynham, J. A. *Biochem. biophys. Res. Commun.* **77**, 714–722 (1977).
- Haslam, R. H. *et al. in Platelets: Cellular Response Mechanisms and their Biological Significance* (eds Rothman, A., Meyer, F. A., Glitter, C. & Silberg, A. 213–231 (Wiley, New York, 1980).
- Blackwell, G. J. *et al. Nature* **287**, 147–149 (1980).
- Hirata, F., Schiffmann, E., Venkatasubramanian, K., Salomon, D. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2533–2536 (1980).
- Cloix, J. F., Colard, O., Rothhut, B. & Russo-Marie, F. *Br. J. Pharmacol.* **79**, 313–321 (1983).
- Di Rosa, M., Flower, R. J., Hirata, F., Parente, L. & Russo-Marie, F. *Prostaglandins* **28**, 441–442 (1984).
- Blackwell, G. J. *et al. Br. J. Pharmacol.* **76**, 185–194 (1982).
- Hirata, F. *J. biol. Chem.* **256**, 7730–7733 (1981).
- Wallner, B. O. *et al. Nature* **320**, 77–81 (1986).
- Hirata, F., Matsuda, K., Notsu, Y., Hattori, T. & Del Carmine, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4717–4721 (1984).
- Mobley, A. & Tai, H. H. *Biochem. biophys. Res. Commun.* **130**, 717–723 (1985).
- Ardlie, N. G., Packham, M. A. & Mustard, J. F. *Br. J. Haematol.* **19**, 7–17 (1970).
- Barber, A. J. & Jamieson, G. A. *J. biol. Chem.* **245**, 6357–6365 (1970).
- Hirata, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3190–3194 (1981).
- Rothhut, B., Russo-Marie, F., Cousin, M. & Lando, D. in *Proc. IUPHAR* Vol. 2, 43–50 (Macmillan, London, 1984).
- Nishizuka, Y. *Nature* **308**, 693–698 (1984).
- Van Den Bosch, H. *Biochim. biophys. Acta* **604**, 191–246 (1980).
- Volpi, M., Molski, T. F. P., Naccache, P. H., Feinstein, M. B. & Sha'afi, R. L. *Biochem. biophys. Res. Commun.* **128**, 594–600 (1985).
- Halenda, S. P., Zavoico, G. B. & Feinstein, M. B. *J. biol. Chem.* **260**, 12484–12491 (1985).
- Habermann, E. & Breithaupt, H. *Toxicol.* **16**, 19–30 (1978).
- Rothhut, B., Russo-Marie, F., Wood, J., Di Rosa, M. & Flower, R. J. *Biochem. biophys. Res. Commun.* **117**, 878–884 (1983).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

Equivalence of intracellular pH of differentiating *Dictyostelium* cell types

David I. Ratner

Department of Biology, Amherst College, Amherst, Massachusetts 01002, USA

During the development of the cellular slime moulds, spore and stalk cells arise from vegetative amoebae. The intracellular pH (pH_i) of the developing cells has been postulated to be important in the choice of morphogenetic programmes¹. This hypothesis stems from the observation that the addition to the culture medium of weak (membrane-permeable) acids or bases favours the differentiation of pre-stalk or pre-spore cells respectively^{1–3}. Similarly, compounds that are known to inhibit a *Dictyostelium* plasma membrane proton pump induce preferential pre-stalk development^{1,4,5}, and a stalk-specific *Dictyostelium* morphogen, 'DIF', has been suggested to act via intracellular acidification^{6,7}. However, in these studies the pH_i values of pre-spore and pre-stalk amoebae were unknown, and the actual effects of the various agents mentioned above on pH_i were also unknown. By monitoring the intracellular fluorescence of a pH-sensitive dye, I have measured the relative intracellular pH of pre-spore compared with pre-stalk amoebae. I report here that, contrary to the above hypothesis, there is no significant difference in the pH_i of the two cell types.

The cytoplasmic pH of vegetative *Dictyostelium discoideum* amoebae has recently been measured as 6.5–6.7 on the basis of ^{31}P -NMR spectroscopy^{8,9} (though the spectra may be open to other interpretations¹⁰). Unfortunately, NMR measurements require more material than can be obtained as purified precursor cell types. In contrast, pH measurements based on the fluorescence of intracellular probes are rapid and highly sensitive. Permeant derivatives of fluorescent dyes readily cross the plasma membrane; negatively charged dye molecules, liberated by intracellular esterases, then accumulate within the cytoplasm. Jamieson *et al.*¹¹ studied changes in pH_i during *Dictyostelium* development by monitoring the fluorescence of cells labelled

Received 25 November 1985; accepted 5 March 1986.

- Knight, D. E., Hallam, T. J. & Scrutton, M. C. *Nature* **296**, 256–257 (1982).
- Rink, T. J., Smith, W. S. & Tsien, R. Y. *FEBS Lett.* **148**, 21–25 (1982).
- Lyons, R. M. & Shaw, J. O. *J. clin. Invest.* **65**, 242–255 (1980).

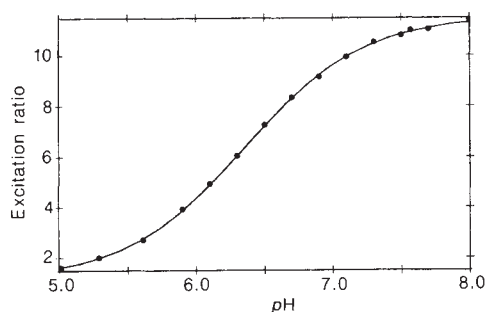


Fig. 1 pH-dependence of the excitation ratio of fluorescein in phosphate buffer. Values plotted are the ratio:fluorescence intensity at 489 nm illumination/fluorescence intensity at 438 nm illumination, as a function of the buffer pH. Experimental values represent the average of two spectral determinations, which were generally found to agree within 1%. The smooth curve is a theoretical fit derived from the observed pK and the fluorescence parameters at the extremes of pH.

Methods. Fluorescein (MCB) was dissolved at $0.2 \mu\text{M}$ in 20 mM phosphate buffers (mixtures of Na_2HPO_4 and KH_2PO_4) of varying pH. pH determinations, using an Orion 601A meter, and fluorescence spectra were obtained at $0-2^\circ\text{C}$. Fluorescence measurements were made with a dual-beam Spex Fluorolog 2 spectrometer controlled by a Datamate, model DM1. Excitation bandwidth 0.9 nm, integration time 1 s, scanning increment 1 nm, emission monochromator set at 510 nm. All spectra were recorded in a ratio mode to correct for the spectrum due to lamp output.

with fluorescein. The following comparison of the pH_i of pre-spore and pre-stalk cells was made using a modification of their method.

The fluorescence excitation spectrum of fluorescein is highly pH-dependent, the major effect being on the relative fluorescence intensities resulting from illumination at the excitation maximum (489 nm) compared with illumination at a shorter wavelength (438 nm)¹¹⁻¹³. Figure 1 shows a standard curve, used to calibrate pH_i measurements, of the excitation ratio of fluorescein in phosphate buffer. Sensitivity is excellent over the pH range 5.8–7.0 (fluorescence $pK=6.4$), a range well suited for examination of *Dictyostelium* amoebae.

When amoebae (vegetative or developing) are incubated with a relatively high concentration ($100 \mu\text{M}$) of fluorescein diacetate, intense but diffuse labelling of the cytoplasm results (ref. 11 and my unpublished results); this indicates that the major contribution to the fluorescence of intracellular dye is cytoplasmic, hence the pH_i reported is expected to be essentially that of the cytoplasm. Figure 2a shows the excitation spectrum of fluorescein-loaded vegetative *D. discoideum* amoebae. Despite its negative charge over the entire physiological range, fluorescein is lost from *Dictyostelium* (and mammalian¹³) cells at an appreciable rate. Therefore, measurements were performed quickly, at low temperature, and correction made for any loss of dye from the cells. Figure 2a includes the spectrum of the cell suspension, that of the cell-free filtrate taken immediately thereafter, and the calculated difference spectrum corresponding to intracellular fluorescein. The pH_i of the vegetative amoebae in Fig. 2a is 6.94. The average pH_i in 14 such measurements was 6.93 ± 0.21 . These measurements were obtained in phosphate buffers spanning 2.5 pH units (pH 5.0–7.5); the intracellular pH was unaffected by such changes in the fluorescence buffer (see also refs 8–10).

The fluorescence procedure is also applicable to pre-spore and pre-stalk cells. The two cell types, obtained from migrating slugs, were purified by gradient centrifugation, and fluorescence

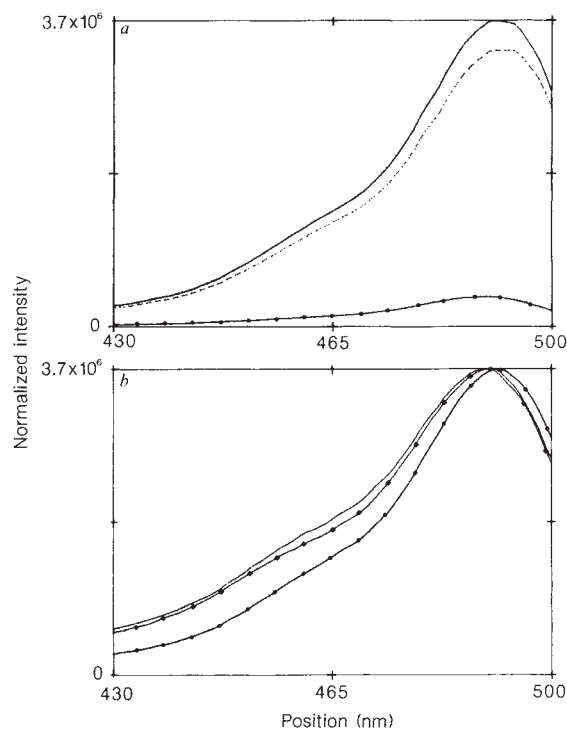


Fig. 2 Fluorescence excitation spectra of vegetative (a) and developing (b) amoebae. a, Spectra for a vegetative cell suspension (—), for the cell-free filtrate (---) dye loss 10%, and for intracellular fluorescein calculated as the difference spectrum of the first two (.....) pH_i 6.94. b, Normalized, filtrate-corrected spectra for three cell types: pre-spore, pH_i 6.08 (—), pre-stalk, pH_i 6.12 (---) and a concurrent vegetative cell preparation, pH_i 6.94 (.....). The pH_i values were determined from the excitation ratio of intracellular fluorescence according to the calibration curve of Fig. 1. Cellular autofluorescence is negligible: the intensity of unlabelled cells is $<10^{-4}$ that of cells containing probe. Note that the difference in pH_i between vegetative and differentiating cells does not result from the choice of strains, as vegetative NC4 amoebae (not shown) have a pH_i equal to that of vegetative AX3 cells.

Methods. Vegetative cells of *D. discoideum* strain AX3 were harvested from axenic medium¹⁸, washed in cold buffer and incubated at $2 \times 10^6 \text{ ml}^{-1}$ in 10 mM PO_4 pH 6.0 (prepared from Na_2HPO_4 and KH_2PO_4), 9 mM NaCl, 9 mM KCl, 2 mM CaCl_2 and $100 \mu\text{M}$ fluorescein diacetate (Sigma) added from a 100-mM stock solution in dimethyl sulphoxide. After incubation at 22°C for 15 min, with shaking, cells were washed three times in cold buffer, the final wash immediately preceding the fluorescence measurement. Pre-spore and pre-stalk amoebae were obtained from migrating slugs of strain NC4, mechanically dissociated in the presence of EDTA, and purified by density gradient centrifugation¹⁹. Labelling of the precursor cells differed from that of vegetative amoebae only in that the incubation period was increased to 25 min. Vegetative amoebae were resuspended at 10^6 ml^{-1} in 20 mM PO_4 pH 7.0; differentiating amoebae were resuspended at $2 \times 10^6 \text{ ml}^{-1}$ in buffer of pH 6.0. Excitation spectra were obtained from 4 ml of cells as in Fig. 1, except that the integration time was reduced to 0.5 s. Immediately after scanning, each cell suspension was filtered through a $0.45\text{-}\mu\text{m}$ membrane filter (Millipore) under gentle suction. The time between the final washing of the suspended cells and the completion of filtration was <4 min. Leakage of dye during this interval was quantitated by determining the relative fluorescence intensities of the filtrate and cell suspension measured at 440 nm. Dye loss generally accounted for 10–15% of the total fluorescence. Suspension buffers were customarily chosen with a pH approximating that of the cells (as revealed in earlier measurements) to ensure that any imprecision in the dye loss correction would have a negligible effect (on the order of 1%) on the calculated intracellular excitation ratio. Because of the small redshift of intracellular fluorescein relative to free dye (the excitation maximum averaging 491 nm rather than 489 nm), the cellular excitation ratio was computed from intensity values at 491 and 440 nm.

measurements were performed as for vegetative amoebae. Figure 2b compares filtrate-corrected excitation spectra for pre-spore, pre-stalk and vegetative amoebae. Although the spectra obtained for the differentiating cells differ from that obtained for the vegetative cells, they are very similar to one another. Indeed, as summarized in Table 1, there is no significant difference in the pH_i of the two precursor populations. As is true of vegetative cells, the measured intracellular pH is independent of that of the surrounding medium (Table 1). However, differentiating

Table 1 Identity of pH_i of pre-spore and pre-stalk amoebae

		Pre-spore		Pre-stalk	
Expt	pH_0	Excitation		Excitation	
		ratio	pH_i	ratio	pH_i
1	6.85	4.78	6.07	5.52	6.20
		3.97	5.92	5.37	6.18
2	6.0	5.65	6.22	5.51	6.20
		4.80	6.08	5.06	6.12
3	6.0	5.47	6.19	5.16	6.14
		5.52	6.20	5.11	6.13
		Mean $pH_i = 6.11 \pm 0.11$		Mean $pH_i = 6.16 \pm 0.04$	
		Mean of all pH_i values = 6.14 ± 0.08			

Experimental methods and data manipulation were as described in Fig. 2 legend. For each of the three experiments shown, pre-spore and pre-stalk spectra were taken alternately. There is no significant difference between the pH_i values of the pre-spore and pre-stalk groups. Similarly, the pH of the cell suspension buffer ($pH_0 = 6.85$ as opposed to 6.0) and the length of time for which cells were held at 0 °C prior to spectroscopy had no effect on pH_i . The excitation ratio is explained in Figs 1 and 2 legends.

slug cells give the somewhat acidic value for pH_i of 6.14 ± 0.08 . The acidification may be the result of accumulation of metabolic acids produced during prolonged slug migration¹⁴, given that the pH_i of amoebae developing in suspension or on pads for shorter periods of time is similar to that of vegetative cells (my unpublished observations; see also refs 8, 10). Because cells within migrating slugs continue to interconvert between pre-spore and pre-stalk (evident immediately in bisected slugs¹⁵ and inferred for undisturbed slugs^{16,17}), the acidic pH_i is, for this reason alone, unlikely to be determinative of the differentiated cell state.

Several controls argue that the fluorescence measurements are a valid indication of relative pH_i values. It is conceivable that the production and accumulation of relatively high concentrations of fluorescein within cells (reaching values in excess of 10^{-4} M; ref. 13 and my unpublished observations) might alter the normal pH_i . However, measurements of the pH_i of vegetative amoebae by the independent criterion of ^{31}P -NMR showed no difference between the pH of highly fluorescent cells and that of control cells ($pH_i = 6.7$ for both; C. Thompson and D.R., unpublished observations). The small disparity between the pH_i of vegetative cells determined by fluorescence (pH_i 6.9) and that determined by NMR spectroscopy (pH_i 6.7) may reflect the particular calibration method used here, or a possible mitochondrial contribution to the fluorescence. With slug cells, varying the degree of labelling suggested that there might be a slight acidification attributable to fluorescein, but the effect was small (<0.2 pH units) and not cell-type-specific (data not shown). Finally, the observed pH_i of total slug cells was found to be identical (pH_i 6.15) whether or not the cells were subjected to gradient centrifugation.

Despite any uncertainty in absolute pH , it is clear that the pH_i values of pre-spore and pre-stalk amoebae are, within experimental error, identical. Thus the hypothesis of a more acidic pH within pre-stalk than pre-spore cells is untenable. This hypothesis has been suggested primarily by the established

effects on cell proportioning of weak acids and bases, and of known or presumed proton pump inhibitors¹⁻⁷. However, recent measurements using both NMR and fluorescence spectroscopy (ref. 10; C. Town, personal communication; and my unpublished observations) have failed to detect the predicted effects of these compounds on pH_i . Rather, the treated amoebae maintain their pH_i ; in so doing, they may perhaps perturb the concentration of some other metabolite(s) whose flux across the plasma membrane is coupled, positively or negatively, to that of protons. Such a metabolite might be the physiological regulator of precursor differentiation in *Dictyostelium*.

This research was supported by NSF grant PCM 8202833. I thank C. Thompson for performing NMR spectroscopy and Dr R. Blankenship for an introduction to fluorescence methods.

Received 24 December 1985; accepted 13 March 1986.

- Gross, J. D., Bradbury, J., Kay, R. R. & Peacy, M. J. *Nature* **303**, 244-245 (1983).
- Town, C. D. *Differentiation* **27**, 29-35 (1984).
- Gross, J. D. *et al. Phil. Trans. R. Soc. B295*, 497-508 (1981).
- Pogge van Strandmann, R., Kay, R. R. & Dufour, J. P. *FEBS Lett.* **175**, 422-442 (1984).
- Serrano, R., Cano, A. & Pestana, A. *Biochim. biophys. Acta* **812**, 553-560 (1985).
- Kay, R. R. & Jermyn, K. A. *Nature* **303**, 242-244 (1983).
- Kopachik, W. J., Dhokia, B. & Kay, R. R. *Differentiation* **28**, 209-216 (1985).
- Jentoft, J. & Town, C. J. *Cell Biol.* **101**, 778-784 (1985).
- Satre, M. & Martin, J. B. *Biochem. biophys. Res. Commun.* **132**, 140-146 (1985).
- Kay, R. R., Gadian, D. G. & Williams, S. R. *J. Cell Sci.* (in the press).
- Jamieson, G. Jr, Frazier, W. & Schlesinger, P. *J. Cell Biol.* **99**, 1883-1887 (1984).
- Ohkuma, S. & Poole, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3327-3331 (1978).
- Thomas, J. A., Kolbeck, P. C. & Langworthy, T. A. in *Intracellular pH: Its Measurement, Regulation and Utilization in Cellular Functions* (eds Nuccitelli, R. & Deamer, D. W.) 105-123 (Liss, New York, 1982).
- Kelly, P. J., Kelleher, J. K. & Wright, B. *Biochem. J.* **184**, 581-588 (1979).
- Gregg, J. & Karp, G. *Expl Cell Res.* **112**, 31-46 (1978).
- Tasaka, M. & Takeuchi, I. *Differentiation* **23**, 184-188 (1983).
- Devine, K. M. & Loomis, W. F. *Dev Biol.* **107**, 364-372 (1984).
- Watts, D. J. & Ashworth, J. M. *Biochem. J.* **119**, 171-174 (1970).
- Ratner, D. I. & Borth, W. *Expl Cell Res.* **143**, 1-13 (1983).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem

☐ Other (please specify) _____
☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
 300 North Zeeb Road, Box 91 Ann Arbor, MI 48106